NATIONAL CONVENTION TRANSACTIONS 1955

NINTH ANNUAL CONVENTION

AMERICAN SOCIETY
FOR
QUALITY CONTROL

NATIONAL CONVENTION TRANSACTIONS 1955

NINTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

MAY 23, 24, 25, 1955
NEW YORK, NEW YORK

PRICE
$3.00 at the Convention
$3.50 if Ordered by Mail

Room 563
50 Church Street
New York 7, New York

Printed in the United States of America By
UNIVERSAL LITHOGRAPHERS, INC.
Baltimore, Md.

FOREWORD

There are different ways of learning about Quality Control. One excellent way is to listen to talented speakers as they present their papers at this Ninth Annual Convention of the American Society for Quality Control. Another way, is to read--and re-read--the written versions of their presentations as contained in these Transactions because it is by such study that their very many worthwhile ideas can be fully understood and appreciated.

There are many misconceptions about what activities may appropriately be considered as constituting "Statistical Quality Control"; many people assert that they have no need for "Statistical Quality Control" but express a sincere interest in statistical analysis of data, the scientific design of experiments, or perhaps a scientific approach to management problems. It will be apparent to anyone who studies the papers contained in these Transactions that leaders in the field of Quality Control make very broad interpretations of the techniques and philosophies to include in this modern science; you will find a multitude of applications representing essentially every area of human interest and endeavor. It is with pride (we hope justifiable) that we present this diversified set of applications of this modern tool of scientific management.

These broad objectives have been achieved because of the excellent planning and energy of the Program Committee in cooperation with the many other Committees of this Convention.

The Program Committee of the Convention Planning Committee has done an outstanding job of scheduling excellent speakers for your listening pleasure at this Convention. Because of the cooperation of these speakers in preparing their talks in advance, these Transactions have been made possible. We appreciate their cooperation, and, in fact, the cooperation of many people in making these Transactions available.

We present them to you with the conviction that you will find many practical and inspirational ideas as you read them.

Ellis R. Ott

Ellis R. Ott
Chairman, Convention Planning Committee

F. Bruce May

F. Bruce May
Chairman, Transactions Committee
Convention Planning Committee

(Foreword continued on next page)

FOREWORD
(continued)

Request for permission to reprint any portion of these Transactions should be addressed to Professor Mason E. Wescott, Chairman, Editorial Board, University College, Rutgers University, New Brunswick, New Jersey.

The paper entitled "Quality Control and Its Application to the Bottling Operation," by William H. von Meyer, pages 167-176 is exempt from copyright restriction and may be reproduced without permission.

Edward M. Schrock
Edward M. Schrock
National Transactions Chairman
General Convention Committee

CONVENTION COMMITTEES
9th ANNUAL A.S.Q.C. CONVENTION

General Chairman
ELLIS R. OTT, University College, Rutgers University

Vice Chairmen
AUGUST B. MUNDEL, Sonotone Corporation
FRANK W. WEHRFRITZ, Esso Standard Oil Company
MASON E. WESCOTT, Rutgers University

Secretary
ALLEN R. CRAWFORD, Rutgers University

Treasurer
JOHN MAJOR, Colgate-Palmolive Company

Arrangements
DAVID VALINSKY, City College of New York
PAUL H. CHOOK, City College of New York

Publicity
DONALD A. TILDEN, Gould and Eberhardt

Transactions
F. BRUCE MAY, Johnson and Johnson

Exhibits
ROBERT R. PERO, Servomechanisms, Inc.

Reception
ROBERT V. WELCH, E. I. duPont deNemours and Company

Program
PHILIP B. PROCTOR, Johns-Manville Corp.

Registration
WILLIAM E. GAUNT, E. R. Squibb and Sons, Division of Olin-Mathieson Chemical Corporation

GENERAL CONVENTION COMMITTEE - A.S.Q.C.

WADE R. WEAVER, Republic Steel Corp., Cleveland, Ohio
FRANK CAPLAN, JR., General Electric Co., Syracuse, N. Y.
CHARLES R. KENDEL, Lamson & Sessions Co., Cleveland, Ohio
PAUL K. LEATHERMAN, Lambert Pharmacal Co., St. Louis, Mo.
BEN H. LLOYD, Leatham and Simpson, Ltd., Montreal, Quebec, Canada
ELLIS R. OTT, Rutgers University, New Brunswick, N. J.
JOHN G. RUTHERFORD, Lear, Incorporated, Grand Rapids, Michigan
ERVIN E. SCHIESEL, The Mattatuck Mfg. Co., Waterbury, Connecticut
EDWARD M. SCHROCK, American Radiator & Standard Sanitary Corp., N.Y.,N.Y.

NATIONAL OFFICERS

PAUL A. ROBERT, President
International Business Machines Corp.

ARTHUR BENDER, JR., Vice-President
Delco Remy Div., G.M.C.

DORIAN SHAININ, Vice-President
Rath and Strong, Inc.

JULIAN H. TOULOUSE, Vice-President
Owens-Illinois Glass Company

E. EUGENE FISHER, Executive Secretary
Bell Telephone Laboratories, Inc.

LEON BASS, Treasurer
General Electric Company

MASON E. WESCOTT, Chairman, Editorial Board
Rutgers University

ALFRED L. DAVIS, JR., Past President
Rochester Institute of Technology

METROPOLITAN SECTION OFFICERS

JOHN MAJOR, Chairman
Colgate-Palmolive Company

ROBERT R. PERO, Vice-Chairman
Servomechanisms, Inc.

DAVID VALINSKY, Vice-Chairman
City College of New York

PAUL H. CHOOK, Secretary
City College of New York

EDWIN SHECTER, Treasurer
Sonotone Corporation

TABLE OF CONTENTS

F. BREWING

G. CHEMICAL

H. ELECTRONICS

L. OPERATIONS RESEARCH

M. STANDARDS

M. TEXTILES

LISTING BY AUTHORS

CONTROL CHART APPLICATIONS IN TEXTILES

Norbert L. Enrick
Institute of Textile Technology

If the production of the average textile mill were to be 100% tested, we would need several thousand times the testing personnel than is required for the mill's regular productive operations. Moreover, since most of the testing involves cutting up, tearing or breaking the material, a program of 100% testing would mean 100% destruction and zero yards of salable merchandise for the mill.

Such a weird prospect does not only force the mill to use sampling as a practical substitute for 100% testing. It also forces the mill to use samples that are exceedingly small portions of the output represented. Often a decision affecting many thousands of pounds must be made from tests on only a few pinches of cotton or short lengths of strand. Statistical evaluation of test results is therefore an important aid in making the right decision. The statistical tools of control charting become one of the choicest methods in the routine supervision of the quality of production. All the typical types of charts have their place, such as for Averages, Ranges, Defects-per-unit and Percent Defective.

From a survey of more than fifty mills in New England, Southern U.S. and Canada, this writer has found charts in successful use in carded cotton spinning applied to the following characteristics:

1. Raw Materials Testing
   The purpose of these tests is to check the quality of raw stock, to allocate it to the proper production lines best suited for a particular end-use, and to properly blend compensating quality characteristics. Tests performed for this requirement, supplemented by control charts, are:

   a. Fiber length in inches.
   b. Length uniformity in Coefficient of Variation.
   c. Fineness in Micrograms per Inch.
   d. Fiber Maturity in Percent.

2. Stock Weight and Variation
   Control of these characteristics will result in a yarn that conforms closely to the desired weight with a minimum of deviation and variability. Charts are generally maintained at each processing stage. These are, in sequence of processing: Opening, Picking, Carding, Drawing, Roving and Spinning. One processing department, usually either the drawing frames or the roving frames, is often used as the key control point for making gear changes to keep weights in line.

3. Linear Uniformity of Stock
   In addition to weight variation, we also need to control the variation in short lengths of a strand of textile material, which occur due to irregular fiber alignment. The laws of chance, as derived from the Poisson distribution, state that the Standard Deviation here cannot be better than the square root of the number of fibers per cross-section of yarn. In practice, the variation will be higher, due to drafting imperfections. For example, a yarn with 100 fibers per average cross-section

would have a theoretically expected Standard Deviation of 10. In actuality, we would normally find a Standard Deviation 30 to 50 percent higher, such as 13 to 15. The degree to which actual variations conform to the theoretical furnishes an indication of how well we have been processing the fibers.

With the development of electronic testing instruments, such as the Brush Uniformity Analyzer with Automatic Evaluator developed by the Institute of Textile Technology, we can now obtain automatic charts that describe the inch-to-inch variations in the textile strand and in addition show on a dial the Average Percent Range (related to Standard Deviation). Thus we have a sort of automatic control chart.

4. Running Conditions of Stock

Running conditions of stock are evaluated in terms of the rate of "ends-down." This refers to the breaking of strands in processing. A broken strand means lost production until it is pieced-up again, and the piecing-up means additional labor cost. Furthermore, a high rate of breaks usually indicates non-uniform, weak, and poorly prepared stock. Ends-down tests are usually performed on the drawing, roving and spinning frames, and expressed in terms of occurrences per thousand spindle hours of regular production. Therefore the defects-per-unit type of chart is used.

5. Reworkable and Non-reworkable Waste

Waste in a cotton mill is of either of the two major classifications above, and then broken down further by department and category within that department. Control charts here aid from the standpoint of quality (proper removal of short fibers, trash, etc.) and cost (keeping avoidable waste to a minimum). Both operator carefulness and proper machine setting may thus be controlled by charting.

6. Processing Tests

A large variety of tests may be included under this heading. The first of these applies to the Degree of Opening and to Feeding Percentage, since improperly opened stock fed at a rate other than optimum is already in violation of one of the prime conditions that make for uniform product. Other tests suited for control chart use are applicable to neps in the card web, to package size of sliver, roving and yarn, to roll settings, to spoon and trumpet knock-off checks, to roving traverse, to critical speed ratios, such as spindle to front roll for proper twist insertion, and to many others.

7. End-Product Evaluation

When the yarn has been spun, it is too late to correct any faults. But nevertheless we like to maintain control charts, so as to assure ourselves that the quality remains up to standard. The characteristics for which mills have kept control charts are these:

a. Single-end strength in grams.
b. Skein-strength in pounds.
c. Linear Uniformity (as discussed above).
d. Defects-per-fifty-yards.

e. Appearance Grade.
f. Twist in turns per inch.
g. Twist variation in Coefficient of Variation.
h. Moisture content in percent.

Proper control of raw materials and processing conditions during production is the best assurance that the final yarn will meet the end-product specifications.

A good measure of the widespread success of quality controls, including control charts, that has been attained in the textile industry is the continuing reduction of doublings of stock in processing. "Doublings" refer to the combining of several strands of material in back of a machine, and then drafting or "attenuating" it to thin out the final combined strand to the thickness of any one of the original strands. In this manner, doubling and drafting is equivalent to the statistical operation of totaling and dividing to obtain averages. Consequently, we may call in the statistical formula for the standard error of sample averages to interpret the effect of doubling and drafting. The formula thus states that the effect is one of reducing variations by the square root of the number of strands combined. In actuality, due to drafting imperfections, the results predicted by the formula will be accomplished approximately but not completely.

For example, in a Canadian cotton mill processing 1-1/16 inch staple, card sliver variation was found to have a Coefficient of Variation of 5.2%. There were 16 doublings in the subsequent drawing operation. Therefore, by dividing 5.2 by the square root of 16, or 4, we obtain 1.3% as the expected Coefficient of Variation for drawing sliver. The actual value was 1.5%.

With the aid of statistical quality control, improved testing equipment, and improved design of machinery, textile mills in the past decade have been making less and less variable stock. This in turn has permitted a constant degrease in the number of doublings in processing at a considerable saving in machinery and labor costs. Thus, while it was common not so many years ago to have several stages of doublings on roving frames and spinning frames, many a mill now prides itself in making a good quality yarn without any doublings at all on these processes. Doublings are still required, however, in earlier processing stages.

CONCLUSION

We may now summarize the principal benefits attainable from statistical quality control in general, and control charts in particular. These benefits are that the control chart aids us in accomplishing the following:

1. More uniform and stronger yarn and cloth, attained through proper allocation of stocks, effective control during processing, and periodic testing of final product.

2. Higher production and lower labor costs, due to decrease of stoppages from broken strands and less labor to piece up strands again.

3. Lower costs due to less doublings of stock.

4. Reduced waste due to effective waste controls.

The degree to which these benefits are attained depends, of course, on the type of stock processed, the end product, the intensity with which statistical controls are used and how well the results are observed—from Management to the front-line production level.

BIBLIOGRAPHY

1. Heard, M. Earl. Research and quality control pay off. Textile Industries 117: 97-108 (March 1953).

2. Grant, E. L. Statistical quality control. New York 36: McGraw Hill Book Co., 1952.

3. Enrick, N. L. Quality control. New York 13: The Industrial Press. 1954.

4. Enrick, N. L. Quality control through statistical methods, Specifically designed for textile mills. New York 16: Modern Textiles Magazine and Rayon Publ. Corp. 1954.

NONRANDOMNESS

John W. W. Sullivan
American Iron and Steel Institute

One aspect of this panel discussion concerns the evaluation of attributes of product units of a lot on the basis of results revealed by a sample of the product units, for the case of nonrandom distribution of attributes. The attributes are characteristics of product units which are evaluated either as conforming or nonconforming with specification requirements.

Another aspect is the method of selecting the product units to obtain the sample, such as by a random method or a nonrandom method. This portion of the discussion concerns methods of selecting the product units that comprise the sample.

This proposal is offered: estimate quantitatively, in advance of selecting the product units, the degree to which a particular method of selecting the units is likely to be (a) "random" or (b) "nonrandom," for the purpose of using that estimate to decide which method of selection shall be used.

It is assumed that the "random" method provides each unit with the same chance of being selected, and the "nonrandom" method does not.

In addition, the proposal is not limited to any manner in which the attributes may be distributed.

On the assumption that the "random" method and "nonrandom" method are mutually exclusive,

$$P_r \text{ plus } P_n \text{ equal one}$$

where P_r is the probability estimate that the method of selecting the product units is random, and
P_n is the probability estimate that the method of selecting the product units is nonrandom.

For a particular estimate of P_n, such as D, the proposal recommends that an arbitrary decision be made to use a nonrandom method of selecting the product units, and thereby avoid the unknown risk of misusing sampling plans based on probability considerations.

In this preprint, the proposal has avoided two questions:

How are P_r and P_n to be evaluated?
What factors govern the evaluation of D?

For correct application of sampling plans based on probability considerations, it is vital that attention be directed first to the practicability of obtaining the sample by a random method of selecting the product units. Otherwise, those plans cannot be used to yield reliable information about the lot.

Is too much being taken for granted in assuming that product units can be selected by a random method without making an estimate of the

probability that the method, in practice, is random?

Consider the following published identifications of samples: random sample, stratified random sample, stratified sample (sometimes called representative or proportional sample), and systematic random sample. Those identifications are described (not necessarily defined) in the Appendix. Regardless of the theory of selecting product units by a random method, is it demonstrable in practice that P_r equals one for those identifications which include the word "random"?

For the purposes of this panel discussion, the two summary questions of this preprint are:

1. Is there a need for estimating the probability that a particular method of selecting product units is random?

2. If the need exists, are means available to evaluate P_r, P_n and D?

Appendix

Random sample. A sample obtained by a selection of items from the population is a random sample if each item in the population has an equal chance of being drawn. Random describes a method of drawing a sample, rather than some resulting property of the sample discoverable after the observance of the sample. (1)

Stratified random sample. If the population to be sampled is first subclassified into several sub-populations, the sample may be drawn by taking random samples from each of the subclasses. The samples need not be proportional to the sub-population size; but, if the combined set of random samples is to be used for purposes of estimating certain population characteristics of the combined population, the assignment of the proportions of the total sample to the sub-populations must be such that

$$\frac{n_1}{N_1\sigma_1^2} = \frac{n_2}{N_2\sigma_2^2} = \ldots = \frac{n_k}{N_k\sigma_k^2}$$

where n_i is the sample size from the ith sub-population with N_i cases and variance of σ_i^2. This sample will be the type for which the parameters may be estimated with minimum variance. (2)

Stratified sample. Let a population be divided into several sub-populations called strata. If from each of these strata random samples are drawn, the resulting pooled sample is a stratified sample. In effect the original population is divided into several sub-populations and random samples are drawn from each. Thus a stratified sample is basically a group of random samples. Let a population be divided into several strata, within each of which: (1) the standard deviation σ_i of the characteristic under analysis is determinable; (2) the frequency is n_i and is known. Then for that system of classification, the stratified sample which provides the minimum variance unbiased estimate of the mean of the characteristic of the population is the one for which the number of random observations for the ith stratum is proportional to $n_i\sigma_i$. If only the n_i are known, then the sampling procedure which minimizes the variance of the estimates of the mean of the population is one in which

the number of observations in the ith stratum is proportional to the n_i. This is sometimes called a representative or proportional sample. (3)

Systematic random sample. Let a population have nk elements, the population being divided into n sub-populations of k elements each. Select a number from 1 to k at random and then sample every kth consecutive element, where 1/k is the ratio of sample to population. This is a special kind of random sample and is in some populations more efficient than simple random sampling. (4)

(1) (2) (3) (4) James, Glenn and Robert C. James (editors): Mathematics Dictionary. D. Van Nostrand Company, New York, 1949, pp. 294, 312.

STATISTICAL CONTROLS APPLIED TO CLERICAL AND ACCOUNTING PROCEDURES

William F. Buhl
Controller's Staff
The B. F. Goodrich Company

Statistical methods, or controls, can be applied to many office and accounting procedures. In fact, they have already been applied by many companies to these functions. They have been applied for the control of clerical error and for obtaining information from clerical work. This information can be for statistical purposes or to determine if error in clerical work justifies checking it 100%. As a result, considerable time can be saved in office operations.

The techniques used in office applications are duplicates of those used by plant personnel in the application of Statistical Quality Control. There are, however, some considerations in office work not present in plant operation:

1. Acceptance - Rectification - We cannot accept or reject in office applications. "Reject"usually implies scrapping which obviously cannot be done. Instead, we accept or rectify. We 100% verify all rejected lots and remove the error.

2. Bivalue Consideration - In office work, particularly where we deal in dollars, we must consider both errors and dollars. Thus, we have two values to contend with. It is not always the question whether it is good or bad - sometimes, we must know how bad it really is. A 1% error in a thousand dollar billing represents a different error than 1% on a hundred thousand dollar billing - although they both represent 1%.

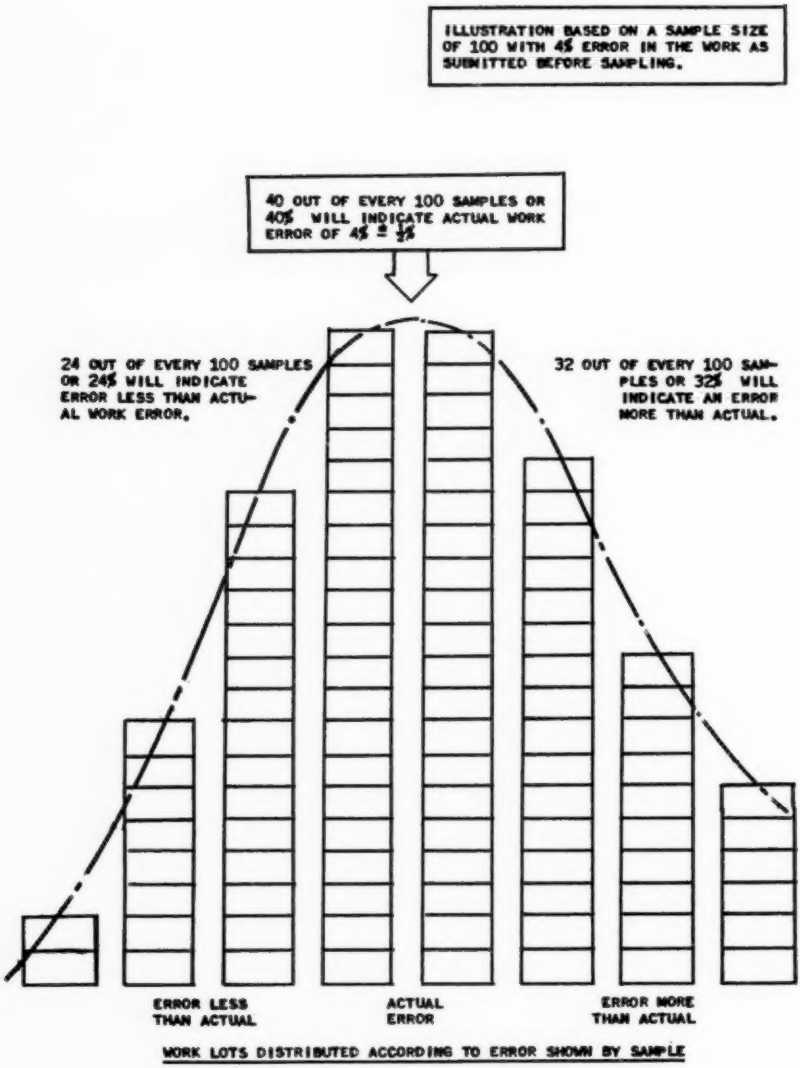
3. Quality Control Substitute - In most offices, the quality of the work is controlled by 100% verification. Installation of a sampling plan is meant to replace this while still obtaining, at least, the same control.

4. Risk Possibility - Office plans must recognize the dollar error possibilities and be designed accordingly. These possibilities represent a greater range in office work (dollars) than they do in plant operations (defectives). We do not have a specific measurement from which we can check variation. An invoice can have any value and the error, likewise, will vary between invoices. Our only common factor is the difference between the correct value of the invoice and the actual value. This represents dollars of error.

In office work as in plant operations, we use sampling to determine the amount of error in a work lot. From this determination, we can decide whether to accept the work as is, or whether complete verification is necessary or desirable. It is possible, therefore, to use this technique in many office and management applications.

To illustrate how sampling operates in actual practice, let us take some specific examples. Let us assume we have a work lot, the error of which is unknown. We will draw a sample at random to determine this rate of error. Our results are determined by the law of probability, which will cause a sample to behave in determinable pattern 997 times out of 1,000.

FIG. - 1 - WHAT SAMPLING WILL INDICATE



Three times in a 1,000, our results may be different, but this difference spread over 1,000 lots will not materially affect our control. We do not obtain this accuracy with present methods.

In drawing samples from a work lot, we do not always find the same error as in the work lot. We do know by the tables of probability what we can expect a sample to show. To illustrate this, let us assume we are drawing a series of samples from one work lot. Each time we draw a sample, we are liable to get a different rate of error in it. We can compute mathematically, or obtain from published tables our chances of obtaining various error rates from a given work lot containing a specific error rate.

Figure 1 shows what we may expect when we sample. For the purpose of illustration, a sample size of 100 with a work error rate of 4% has been selected. Under these conditions, based on the law of probability, we can expect that 40 out of every 100 lots will show the actual error of 4%. The balance of the samples will be distributed approximately evenly over and under the actual error. In this particular illustration, 32 samples will show more than 4% and 24 will show less than 4%.

This type of distribution, referred to as the normal distribution, will occur each time you sample. The shape of the curve will change depending on the size of the sample and the rate of the error in the work. An increase in the sample size will decrease the spread of the base of the curve. Less variation will occur. Large sample sizes will have a small variation. A decrease in error rate obviously will also result in a narrower curve because there is less error to vary.

It is possible, therefore, to determine in advance just how much a sample will deviate from the actual, and from such knowledge, develop a sampling procedure. If the sample indicates an error rate in excess of the computed variation, it is an indication that the error has changed.

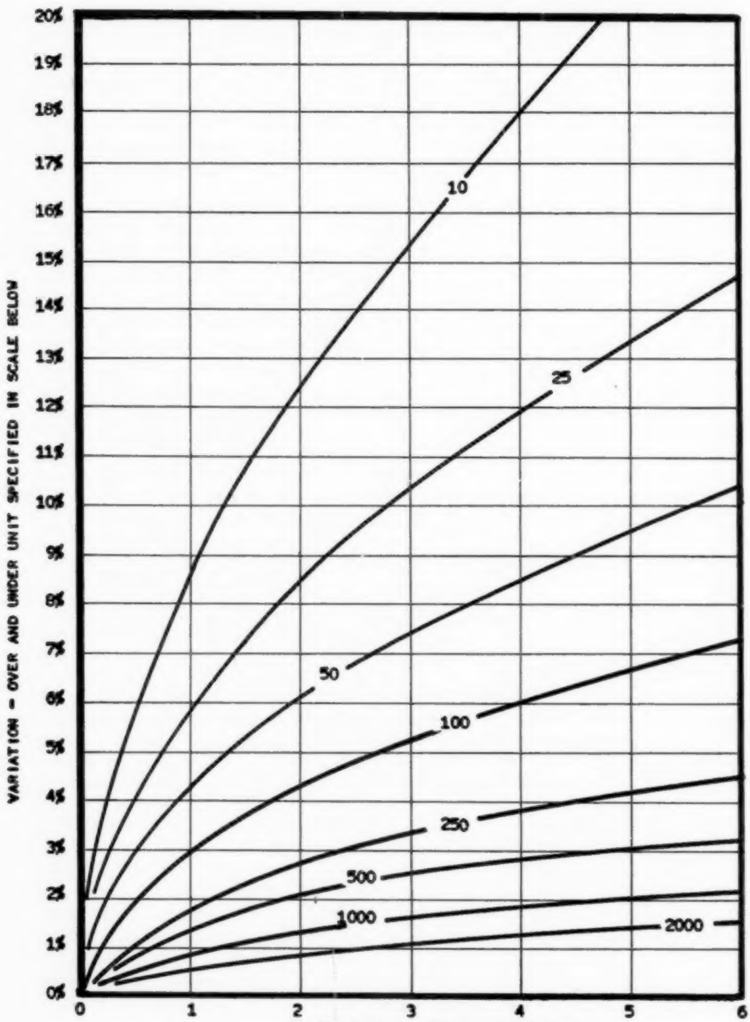
The application of SQC can best be illustrated by taking one of the more common office routines. Let us take the routine of checking invoices. The question constantly before us is whether the work contains error. Three courses have been our choice in the past. We could take a chance; we could spot check the work; or the most common solution, check them all. We now know that each of these methods has some disadvantages.

1. Take a Chance - This has a high element of risk. There is no way of telling how much error is in the work.

2. Spot Check - This is often confused with sampling. The two have only one thing in common - the selection of a portion of the work for checking. Spot checking does not insure randomness (each piece of paper having an equal opportunity of being selected). It is also possible to spot check too many invoices for a satisfactory control - or not spot check enough. The risk of bias (selecting by color, value, etc.), and the inability to determine when a sufficient amount has been checked, makes its use rather hazardous.

3. 100% Verification - This requires more work than is necessary to provide control - also, it is not 100% accurate. It does provide, however, a satisfactory control with a minimum of risk but at a high cost.

FIG. - 2 - SAMPLE VARIATION



EXAMPLES:- SAMPLE OF 100 WITH 1% ERROR IN WORK. VARIATION + AND - 3%.
ANSWER - SAMPLES WILL INDICATE BETWEEN 0% AND 4% ERROR.

SAMPLE OF 1,000 WITH 5% ERROR IN WORK. VARIATION + AND - 2%.
ANSWER - SAMPLES WILL INDICATE BETWEEN 3% AND 7% ERROR.

Random sampling with its computed risks provides greater accuracy with less effort, improves recovery with less fatigue and monotony, and results in a lower error rate. No other method, known today, provides all these advantages. Let us consider each one briefly:

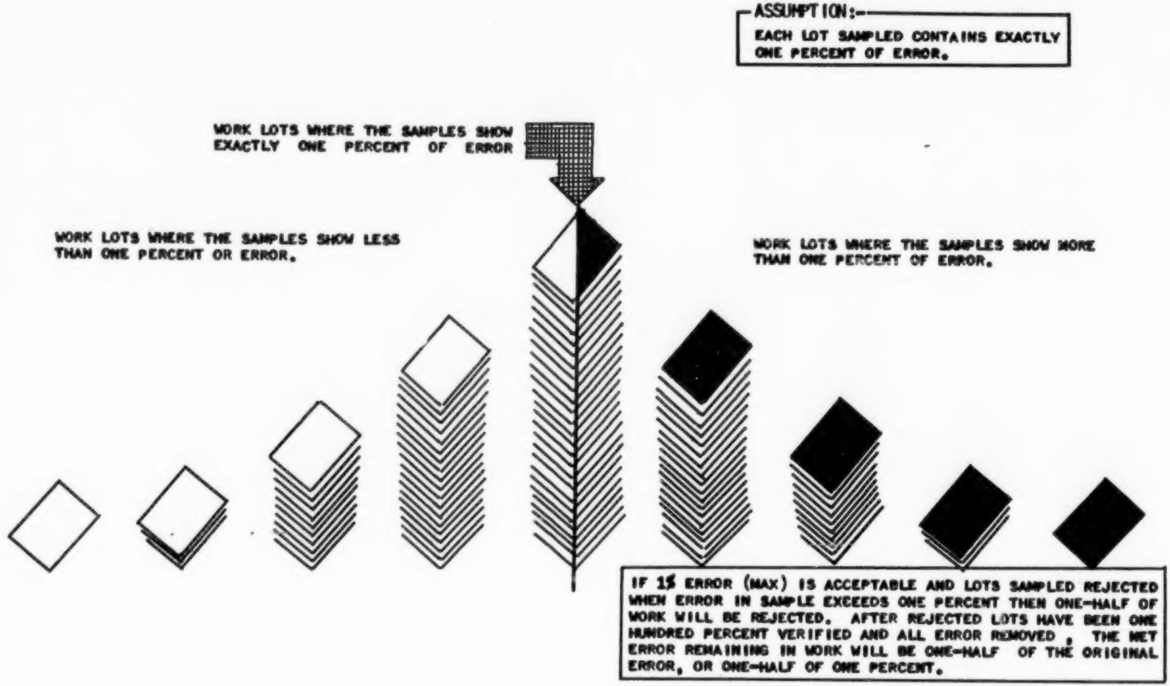
Greater Accuracy	Sampling permits a fast and reliable means to determine error in clerical work - permitting immediate corrective action. Concentration on those lots containing error promotes more efficient checking.
Less Effort	Only those lots which require it are checked. If sample does not show sufficient error, the lots are accepted without further checking.
Improved Recovery	By the concentrated check on lots containing error, a better checking job can be performed - result more recovery.
Less Fatigue and Monotony	One hundred percent verification results in fatigue and monotony by checking large quantities of work without finding error. Sampling provides lots to check which contain error.
Low Rate of Error	Companies installing control charts in connection with sampling, report as high as a 74% reduction in error. Chart provides a visible means of showing an employee her error - and it encourages error reduction.

It can be seen, therefore, that there are many advantages to be obtained from sampling in place of examining every piece of paper. The question arises, however, as to how we can have confidence in a sample when it was shown in Figure 1 that the samples will not always show the actual - in fact considerable variation will exist. The fact is that we can have confidence in sampling because we can determine this variation. We know in advance what variation we may experience with a given work error, and can therefore determine when the error rate has changed. The extent of this sample variation is shown in Figure 2.

The curved lines in this illustration represent various sample sizes. The scale at the bottom is the percent of error in the work lot or permitted to be in work lot. The vertical scale is the percent of error to be added and subtracted from the actual error average to obtain the total range of error the sample will show. It may be mentioned here that the scales can also be used to represent defects per hundred or dollars per hundred if control is desired on those items.

If 2% error is in work lot (or it is desired to control error to 2%), it can be seen from the graph that a sample of 100 will vary on the average 4.2%. This means that the samples may run as high as 6.2% and as low as zero. Actually the lower point will be minus 2.2% but we cannot obtain less than no error. Increasing the sample to 250, lowers the variation to 2.8%. A 1,000 sample will result in 1½% variation or show from ½% to 3½% error. Thus, it can be seen that we can predict what will happen on the average. We know that if our sample of 1,000 shows between ½% and 3% error, there is a strong probability that it was drawn from a lot containing 2% error.

FIG. - 3 - HOW SAMPLING CONTROLS THE ERROR REMAINING IN WORK AFTER VERIFICATION OF REJECTED LOTS.



The variation of a sample is therefore the key to the entire sampling picture. If we can predict with reasonable certainty how a sample will behave under specific conditions, then there is no need for us to look at all of the work to obtain the information we desire.

Where the sample variation results in a minus quantity, then another consideration may be necessary. This occurs when the variation is more than the average. For example, if the variation is 4% and the average error 2%, the lower limit of the variation will be -2%. In such cases a large number of samples will show no error. While this will not alter our ability to control, it may have an effect on any observations we may desire to make. In such cases, a sample size can be selected to avoid a variation which results in a minus quantity. Table 1 below shows the probability of no error appearing in a sample under various conditions.

Table 1

Size of Sample	Probability of No Error Appearing in the Sample if the Percent of Error in Work Submitted is - 1	2	3	4	5	6	7
10	90%	82%	74%	67%	61%	55%	50%
25	78%	61%	47%	37%	29%	22%	17%
30	74%	55%	41%	30%	22%	16%	12%
50	61%	37%	22%	14%	8%	5%	3%
100	37%	14%	5%	2%	0%	0%	0%
250	9%	0%	0%	0%	0%	0%	0%

Up to this point we have been talking about taking a series of samples from a <u>single</u> work lot. In actual practice we make our decision from one sample drawn from a work lot. Figure 3 shows us what is likely to happen under those conditions. In this illustration, we are assuming an error rate of 1% in the work and drawing a specific sample. A sample drawn from this lot may fall in any one of the piles. Forty percent of the time it will show the actual error. The balance of the time it may appear on either side. Let us further assume that 1% error is the maximum we will permit in the work. Not having any other standard at this point, we can only decide that if the sample shows more than 1% error, we will reject the work to be 100% verified. Obviously this standard is incorrect as can be seen from the chart. If those samples which show more than 1% are rejected, it will result in our 100% checking half of the work. Removing all error from one-half of the work containing 1% error will result in ½% error remaining. A standard or control point is required to determine at which point work should be rejected. Offhand, it would appear to be 2% since it will be reduced one-half. Actually the control point is determined by the size of the sample, the error rate, and the tables of probability.

In actual practice, the net error remaining in the work will be slightly higher than mentioned because 70% of the lots will be accepted. The 70% represents the 30% showing less than the actual error and 40% showing the actual error. Only 30% would be rejected for 100% verification.

To illustrate how probability tables are used to set control points or acceptance limits, let us take a case where the work lot contains 5% error, which is permissible, and the sample size is 100. On this basis, the average errors which will appear in the sample will be 5% of 100 or 5. We can now consult the probability tables and refer to an average occurrence of 5.

Table 2

Average Occurrence	Probability of Occurrence of: 0 Error	2 or Less	4 or Less	6 or Less	8 or Less	10 or Less	12 or Less	14 or Less
5	0.7%	12.5%	44.0%	76.2%	93.6%	98.6%	99.8%	100%

The probability table (Table 2) informs us that we can expect 14 or less errors to appear in the sample. In setting our control points, however, we do not use 100% probability as the governing point. We use 95% to give us full assurance that the average error remaining in the work will not exceed 5%. If we used 100%, the "border line" readings we obtain in sampling will result in an average higher than 5% (9.4%) as can be seen in the table below. Using 95% probability, we find that slightly more than 8 errors would be the maximum average occurrence. This then becomes our acceptance limit. If a sample contains 9 or more errors we will reject the work for a 100% verification. The net result will limit the error in the work to 5.1% (Table 3).

Up to this point, we have been talking about specific error rates in the work. What happens when the error rate changes? If the error rate decreases, there are fewer lots rejected. As the error rate increases, more and more lots are rejected for 100% verification resulting in a lower net error remaining in the work. Table 3 below shows how sampling functions under these conditions.

Table 3

% Error in Work Submitted	95% Probability: % Lots Accepted	95% Probability: Net Error in Work	100% Prob. Table: % Lots Accepted	100% Prob. Table: Net Error in Work
5%	93.2	4.7	100	5.0
6%	84.7	5.1	100	6.0
7%	72.9	5.1	99.4	7.0
8%	59.3	4.7	98.3	7.9
9%	45.6	4.1	95.9	8.6
10%	33.3	3.3	91.7	9.2
11%	23.2	2.6	85.4	9.4
12%	12.5	1.9	77.2	9.2
13%	10.0	1.3	67.5	8.8
14%	6.2	0.9	57.0	8.0
15%	3.7	0.6	46.6	7.0

It can be seen that the maximum average error remaining in the work will be 5.1% when the error in the work submitted is 6% to 7%. As the error in the work increases, the error remaining will decrease. When the

error in the work reaches 23%, all lots will be rejected - thus resulting in no error after verification. This table also illustrates why we do not use a 100% probability basis. If we did use such basis (acceptance limit 14 errors), the net error remaining in the work would reach a maximum average of 9.4%. Using the 95% probability basis (acceptance limit 8 errors) we are assured of a maximum of 5.1% error as desired. An actual experience using these principles was obtained in a field check recently made by us.

> To illustrate these principles to our field auditing staff we ran a test on checking inventory cards. Two hundred and sixty-four entries were examined 100% and 16 errors were located or 5.8%. The next step was to prove to the auditors that we could obtain the same results by sampling. Every tenth entry was compared which resulted in a sample of 27. As the error in the work was 5.8%, the average sample error should be 5.8% of 27 or 1.6. Consulting the 95% probability basis for an average of 1.6, we found that we should not obtain more than an average of 3 errors.
>
> The sample of 27 was taken and exactly 3 errors appeared in the sample, proving that the average error was 5.8% with only 1/10 of the work effort. In actual operation, the auditor at this point would have to decide whether the error indicated in the work lot was sufficient to warrant a 100% audit.

The methods illustrated here have been used to develop sampling plans without the mathematical computation and formula common to Statistical Quality Control. Anyone interested in the statistical theory and the mathematical background of these techniques can readily find them in any of the excellent textbooks on Statistical Quality Control. The probability tables shown were taken from the text of the book on Statistical Quality Control by Grant.

By using this method, you can easily develop your own sampling plans for the simpler applications. It will control the average error remaining in the work after verification of the rejected lots. All that is needed in addition to the probability tables is some idea as to the proper sample size. Listed below in Table 4 are the recommended sample sizes for various work lots taken from MIL-STD 105A published by the Government Printing Office.

Table 4

Lot Size	Sample	Lot Size	Sample	Lot Size	Sample
2-8	2	66-110	15	801-1300	110
9-15	3	110-180	25	1301-3200	150
16-25	5	181-300	35	3201-8000	225
26-40	7	301-500	50	8001-22000	300
41-65	10	501-800	75	22001-110000	450

In comparing this table with the audit test previously referred to it can be seen that the sample of 27 is slightly less than the 35 recommended for a work lot of 264. This emphasizes the rule that sampling should not be done on a percentage basis.

There is one further consideration to be made in designing sampling

FIG. 4---AUDIT OF CUSTOMER'S INVOICES

LOT SIZE: 1,000 INVOICES - SAMPLE SIZE: 100 - ACCEPTANCE LIMIT: $2. ERROR

RESULTS OF SAMPLING

LOT NO.	ERROR LOCATED SAMPLE DOLLARS	ERRORS LOCATED ON 100% VERIFICATION	
		DOLLARS	QUANTITY
1	$ 5.36	$ 39.54	6
2	NONE	NOT VERIFIED	
3	24.26	60.33	8
4	12.65	98.99	9
5	NONE	NOT VERIFIED	
6	NONE	NOT VERIFIED	
7	NONE	NOT VERIFIED	
8	1.19	NOT VERIFIED	
9	NONE	NOT VERIFIED	
10	39.86	137.69	14
11	NONE	NOT VERIFIED	
12	NONE	NOT VERIFIED	
13	NONE	NOT VERIFIED	
14	NONE	NOT VERIFIED	
15	NONE	NOT VERIFIED	
16	20.58	330.16	33
17	NONE	NOT VERIFIED	
18	NONE	NOT VERIFIED	

LOT NO.	ERROR LOCATED SAMPLE DOLLARS	ERRORS LOCATED ON 100% VERIFICATION	
		DOLLARS	QUANTITY
19	$ 2.00	NOT VERIFIED	
20	1.50	NOT VERIFIED	
21	94.35	$118.63	8
22	NONE	NOT VERIFIED	
23	NONE	NOT VERIFIED	
24	NONE	NOT VERIFIED	
25	NONE	NOT VERIFIED	
26	NONE	NOT VERIFIED	
27	NONE	NOT VERIFIED	
28	NONE	NOT VERIFIED	
29	NONE	NOT VERIFIED	
30	14.50	53.23	7
31	NONE	NOT VERIFIED	
32	NONE	NOT VERIFIED	
33	NONE	NOT VERIFIED	
34	1.35	NOT VERIFIED	
35	4.60	36.46	14
TOT	$222.20	$875.03	99

COMPARISON

EXPLANATION	LOTS AS LISTED IN TABLE ABOVE	RESULTS AFTER -3-MONTHS OPERATION
TOTAL - ALL INVOICES (1,000 LOTS)	35,000	113,000
TOTAL - INVOICES VERIFIED (SAMPLES AND REJECTED)	10,700	38,694
PERCENTAGE OF WORK CHECKED	30.6%	34.2%
NET WORK REDUCTION IN PERCENT	69.4%	65.8%
NUMBER OF LOTS SAMPLED	35	154
NUMBER OF LOTS REJECTED	8	43
PERCENTAGE OF LOTS REJECTED	22.9%	27.9%
ERRORS LOCATED - TOTAL	99	246
PERCENT ERROR TO TOTAL INVOICES	0.283%	0.218%

PREVIOUS ERROR PERCENTAGE ON 100% VERIFICATION OF ALL INVOICES - - 0.222%

plans for office application - the control of dollar error. We have conducted many tests and surveys on the distribution of dollar error and have found that, in general, they conform to normal distributions. In controlling such dollar error, it is necessary that we convert to another basis other than one of percentage. Percentage of error may not always be satisfactory because of the vast difference between 1% error on a $1,000 billing and a 1% error on a $10 billing.

After many tests, we have adopted a dollar error per invoice plan. This is an adaptation of the defects per unit method explained in SQC literature. We have modified it to allow one dollar in error to represent one defect. This makes it possible to use the percentage defective techniques by considering them as defects per 100 invoices or as modified - dollar error per hundred invoices.

To use this plan, it is necessary to survey past errors, or to determine the permissible dollar error per hundred invoices. After this has been determined, the methods previously outlined are used. For example, referring to Tables 2 and 3 - 95% probability - if the permissible error is $5 per hundred invoices, then our acceptance limit will be $8 per hundred and the maximum average error left in work $5.10 per hundred invoices. If the sample size is 150, it will be necessary to convert the acceptance amount to $12 for this sample.

Figure 4 shows the results of an installation we have had in operation for about five months on the auditing of Field Billing. The upper table shows the results of 35 days sampling. The work lots contain 1,000 invoices and a sample of 100 is taken. If the error exceeds $2, the lot is rejected for 100% verification. In the 35 lots shown, only 8 were rejected. It should be noted that each rejected lot recovered more than $20 ($2 per hundred) which is what the plan was designed to do.

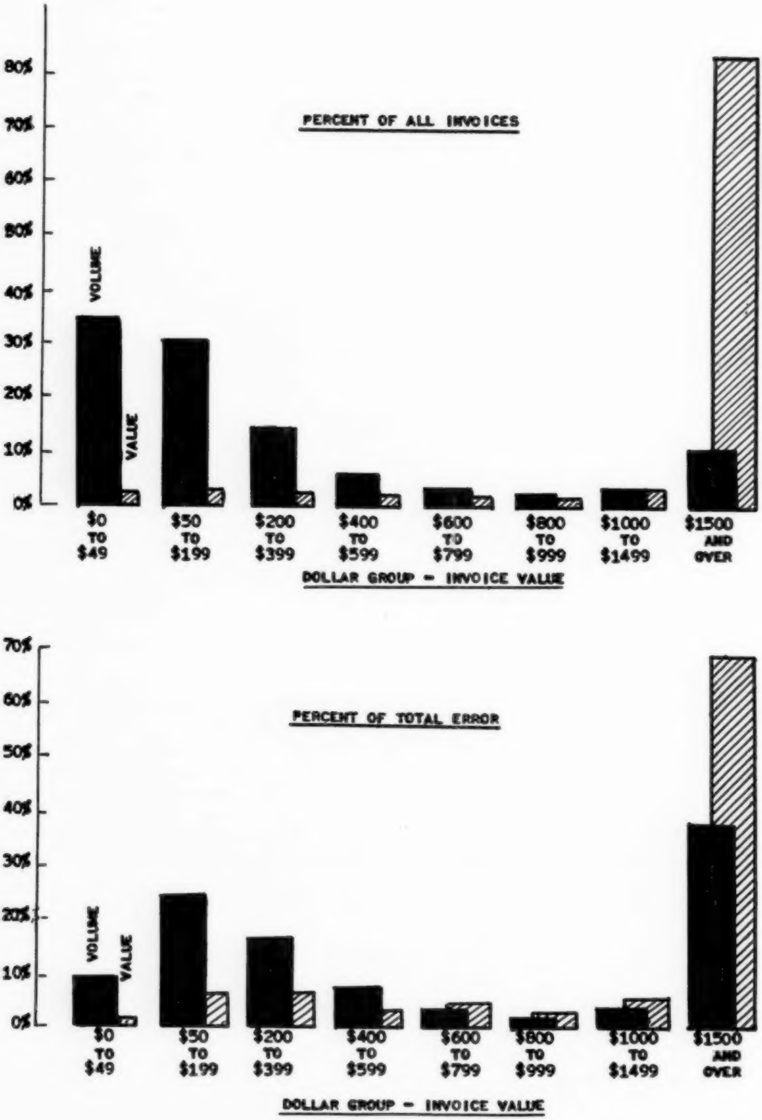
The lower portion gives a comparison between the 35 lots and the 154 lots taken over a three month period. It is interesting to note that the average error found by sampling was 0.218% compared to 0.222% under the previous method with 65.8% less work.

This field of auditing is a "natural" for sampling techniques. Most of the audit programs performed by public accounting firms or by industry include checking of various types. Much of this is done on a spot check basis. As previously mentioned, this spot checking does not contain the assurances present with sampling. For example, one common method of auditing is to select three arbitrary periods - one at the beginning of the audit program - one at the end - and one somewhere in between. With this method the auditor only has a 25% assurance that he will find something over a twelve month audit. Errors can appear in nine of the months and not be detected unless other indications are apparent.

Under a sampling plan, however, a small sample can be taken over the twelve month period to indicate to the auditor where the check should be concentrated. Another sample could be taken of this period to determine whether a 100% audit is warranted.

Audit applications can be made to verification of accounts receivable, checking of stock ledger cards, checking of endorsements, verifying all types of supporting documents, and many other applications too numerous to outline here.

FIG. 5 - DISTRIBUTION OF ERROR IN VENDORS INVOICES



Another method of controlling dollar error is by "stratification". Here invoices are divided into dollar groups and a different plan applied to each group. High value invoices would be 100% verified. Sampling of the lower values would be in proportion to the value. To select a plan of this type, a detailed analysis must be made of the value and volume of invoices handled and the extent of error.

A survey of this type is necessary to obtain an indication of the distribution of the dollar value and the extent and location of the error both as to dollar value and quantity. From this data, it will then be possible to determine at which point it would be more economical to check all invoices because of the smaller volume involved and the greater risk due to high value errors. Such stratification can also be carried into greater detail by subdividing the balance into various price groups. If this is done, smaller sample sizes can be taken as the risk decreases. In fact, the survey may indicate that the low value invoices, usually representing a large volume, need not be checked at all.

Figure 5 shows the results of such a survey on vendors' invoices. It can be seen that only 11% of the volume but 84% of the value are represented by invoices over $1,500. These invoices also contain 37% of the errors representing 68% of the total recovery. By 100% verifying invoices over $1,500, we are only handling 11% of the paperwork, but are controlling 68% of the dollars recovered by the corporation. The remainder of the invoices ($50 to $1,499) can be sampled. In addition, a further simplification can be made by eliminating the ehecking of invoices under $50. These represent 34% of the volume but only 1% of the recovery. A sample could be taken of this group every third month to determine if the pattern has changed.

A review of Figure 5 will also indicate that the intermediate values could possibly be stratified into two or more groups for a sampling advantage. Final decision will be determined by the volume involved and the cost of sorting the invoices as against the benefits gained.

Up to this point, we have been illustrating sampling plans designed to indicate when the error in the work is sufficient to warrant 100% inspection. This is not the only use to which statistical methods can be applied to office and accounting procedures. Another use which opens a much wider field is the taking of samples to obtain averages - averages which can be used in a multitude of computations.

There are many gains in using these techniques for that purpose. First of all, we can determine the sample size required for any desired degree of accuracy. This gives us assurance that our average will be comparable to that obtained by examining every piece of paper. We gain by securing our information in less time and with less chance for error.

The averages obtained by sampling can be compared and by use of statistical techniques, it can be determined if the changes are significant, i.e. whether they require any investigation to determine the cause of any definite variation. If not, no further action is required, resulting in a considerable saving in time.

It is obvious that statistical methods provide a potent tool in every day clerical, accounting and other office applications. A tool, if fully utilized, can result in considerable savings of time, money and also provide better management control.

FIG. 6--COSTING BY AVERAGES

COMMODITY #1	COMPUTED COST		COST BY SAMPLING AVERAGES			NET DIFFERENCE	
	ITEM QUAN	COST AS COMPUTED	ITEM QUAN	COMPUTED SQC COST	CHANGE IN %	CALCULATION	DOLLAR CHANGE
FIRST MONTH	18	$ 5,908	1	$ 5,917	100.2	- 17	+ 9
SECOND MONTH	15	$ 4,837	1	$ 4,855	100.4	- 14	+ 18
THIRD MONTH	17	$ 4,625	1	$ 4,592	99.3	- 16	-33
TOTALS	50	$ 15,370	3	$ 15,364	99.96	- 47	- 6
COMMODITY #2							
FIRST MONTH	15	$ 5,586	1	$ 5,504	98.5	- 14	+ 82
SECOND MONTH	20	$ 12,660	1	$ 12,687	100.8	- 19	+ 27
THIRD MONTH	19	$ 14,256	1	$ 14,227	99.8	- 18	- 29
TOTALS	54	$ 32.502	3	$ 32,418	99.74	- 51	- 84
COMMODITY #3							
FIRST MONTH	15	$ 7,776	1	$ 7,828	100.7	- 14	+ 52
SECOND MONTH	17	$ 5,823	1	$ 5,882	101.0	- 16	+ 59
THIRD MONTH	22	$ 15,025	1	$ 14,899	99.1	- 21	-126
TOTALS	54	$ 28,624	3	$ 28,609	99.95	- 51	- 15
ALL COMMODITIES							
FIRST MONTH	48	$ 19,270	3	$ 19,249	99.9	- 45	- 21
SECOND MONTH	52	$ 23,320	3	$ 23,424	100.5	- 49	+104
THIRD MONTH	58	$ 33,906	3	$ 33,718	99.4	- 55	-188
TOTALS	158	$ 76,496	9	$ 76,391	99.86	-149	-105

PROCEDURE USED:- SELECT SAMPLE OF MONTH'S PRODUCTION BY PREDETERMINED COMMODITY GROUP. COST SAMPLE BY SUCH GROUP. OBTAIN AVERAGE COST. MULTIPLY TOTAL OF PRODUCTION IN EACH COMMODITY GROUP BY THE SAMPLE AVERAGE OF SUCH GROUP.

EXPLANATION AND CONCLUSIONS:- COSTING BY AVERAGES RESULTS IN 94% LESS COMPUTATIONS WHILE OBTAINING THE SAME RESULTS WITHIN AN AVERAGE OF 14/100 OF 1%.

IT IS DOUBTFUL WHETHER THIS ACCURACY IS BEING MAINTAINED AT PRESENT WITH THE MULTITUDE OF COMPUTATIONS INVOLVED.

Various organizations throughout the country have made use of these techniques in their operations, for the benefit of the office and to the management. In our organization, we have been experimenting with the use of the techniques in various applications. One of these is the use of averages to cost our production as shown in Figure 6.

> In the test shown, three commodity groups were analyzed for a three month period and the number of computations required and the dollar cost obtained recorded. The unit cost of the items produced in any month were then averaged. This average was then multiplied by the total production for the month for that commodity group. The second group of columns shows the result of this method. Observe the accuracy obtained - 99.74 to 99.96%. The differences in the monthly totals are not large - varying a maximum of $126. This accuracy was obtained with three computations in place of an average of 50 - a 94% reduction in effort. It is doubtful whether the present computing method can furnish such accuracy.

This application opens a wide field in office applications where a large volume of repetitive computations are performed. By sampling for an average, and then using such average for computing, considerable time can be saved. Statistical Quality Control techniques are employed to determine the sample size, the frequency of sampling, and the maximum variation possible or desirable.

United Air Lines is currently conducting an experiment along these same lines by using the technique to bill other air lines for passengers carried by United. Minnesota Mining and Manufacturing use the technique for forecasting inventory requirements.

Two other uses of this technique have been made by us. Both of these were one time applications, but they serve to illustrate how the technique can be applied.

In one instance, one of the procedure men was having difficulty in securing approval of a new invoice form which provided for a smaller number of items per invoice. The objection to the new form was that too many sheets would be required per billing. To overcome this objection, a sampling or one year's billings (approximately 5%) was taken to obtain an average within a possible error of one item. The results proved that the average was less than the number provided by the new form and it was accepted.

In another instance our Accounts Payable Department was interested in determining the number of one item purchase orders that the corporation issues on the average. Computation of the sample size with a possible error of 1% indicated that 300 should be observed. Observation of this sample revealed that 60.4% of the purchase orders contained one item. Skeptical of the results, the department checked 6,000 purchase orders and obtained an average of 60.0%.

Similar sampling programs have been used to determine the average claim per accident, and to audit a representative number of claims.

The techniques can be used on any application where it is necessary to obtain information from clerical work. By using these tools, we can determine the sample size required for any degree of accuracy.

FIG. 7--BILLING BY AVERAGES

RECORD OF ACTUAL BILLINGS IN EACH OF THE SEVEN INTERVALS LISTED.

PRICING A 20% SAMPLE FROM EACH MONTH'S INVOICES AND OBTAINING AN AVERAGE COST FOR SUCH SPECIFIC MONTH.

TOTAL BILLING AVERAGE METHOD

VARIATION BETWEEN METHODS DOLLARS

AMOUNT OF BILLING WHEN USING A THREE MONTH AVERAGE USED FOR EACH MONTH

PERIOD	ACTUAL BILLINGS		RANDOM SAMPLE OF BILLINGS			BILLING BY AVERAGES		FOUR MONTH AVERAGE	
	TOTAL SHIPPED	BILLING DOLLARS	TOTAL UNITS	DOLLAR VALUES	AVERAGE VALUE	BILLING DOLLARS	VARIAT'N IN DOLLS.	BILLING DOLLARS	VARIAT'N IN DOLLS.
FIRST MONTH	1,541	$ 40,437	347	$ 7,140	$20.58	$ 31,714	-8,723	$42,193	+1,756
SECOND MONTH	1,542	45,484	253	7,507	29.67	45,751	+ 267	42,220	-3,264
THIRD MONTH	1,653	45,388	342	9,240	27.02	44,664	- 724	45,259	- 129
FOURTH MONTH	1,567	42,001	228	7,656	33.58	52,620	+10,619	42,904	+ 903
FIFTH MONTH	1,757	47,787	409	10,448	25.55	44,891	-2,896	48,107	+ 320
SIXTH MONTH	1,190	28,991	335	7,450	22.23	26,454	-2,537	32.582	+3,591
SEVENTH MONTH	1,841	53,451	312	8,582	27.51	50,646	-2,805	50,407	-3,044
TOTAL 7 MONTHS	11,091	$303,539	2,226	$58,023	——	$296,740	-6,799	$303,672	+ 134
TOTAL 4 MONTHS	4,609	$126,263	451	$12,346	27.38	$126,194	- 69	——	——

EXPLANATION AND CONCLUSIONS - - THE WIDE VARIATION ($10,619) RESULTING FROM BILLING ON AN AVERAGE BASIS MONTHLY, IS DUE TO THE SMALL SAMPLE SIZES INVOLVED. THE 20% SELECTION USED HERE FURNISHED ONLY BETWEEN 4 AND 12 INVOICES IN A SAMPLE - A SIZE TOO SMALL TO INSURE A SMALL VARIATION. AS A RESULT, THE TOTAL VARIATION AMOUNTS TO $6,799.00. WHEN USING A FOUR MONTH SAMPLE (35 INVOICES) THIS DIFFERENCE FOR A FOUR MONTH PERIOD WAS REDUCED TO $69.00. BY USING THIS AVERAGE ($27.38) AND APPLYING IT TO EACH INDIVIDUAL MONTH THE VARIATION FOR SEVEN MONTHS AMOUNTED TO ONLY $134.00 ON OVER $300,000. BILLING.

IT CAN BE CONCLUDED THAT THIS METHOD IS FEASIBLE PROVIDING AMPLE SAMPLE SIZES CAN BE SECURED WITHOUT MAKING A 100% VERIFICATION AND PROVIDED ALL FACTORS CAN BE REFLECTED IN METHOD OF COMPUTATION.

A survey is now under way for another application of a similar nature. This application refers to inter-company billings of finished goods and is shown in Figure 7.

> The results shown here indicate that there is a basis for this method of billing although additional studies are required to remove the extreme dollar variation. The method used was to select a sample of the billings to be priced and price only the sample. An average price was then obtained and multiplied by the total unit shipped each month.
>
> The extreme dollar variation on a monthly basis is due to the small sample sizes selected. On a seven month basis, the dollar billing comes within 2.2% of the correct amount. In another test of the same commodity on a four month basis, the variation was reduced to $69. Using this average obtained on a four month basis and applying it to the shipments for the seven months, the final answer comes within $134 in place of $6,799. This seems to indicate that it may be possible to take a prior four month average and use it for a subsequent period. Definite conclusions cannot be stated from these preliminary tests. They are reproduced here to show the type of applications possible.

Only a few of the many possibilities of using SQC techniques in clerical and accounting procedures have been shown here. You, too, will find these techniques helpful in reducing your work verification or in obtaining information from the work. Many installations are possible in the office without involved mathematics, statistics, or theory. Truly, this background is helpful in solving the more complex applications of SQC. For the purposes of the office manager or the procedure analyst, the applications can be rather simple and yet effective.

TRUNCATED AND CENSORED SAMPLES FROM NORMAL DISTRIBUTIONS*

A. Clifford Cohen, Jr.
The University of Georgia

1. INTRODUCTION

Samples obtained when selection or observation is restricted over some portion of the range of possible population values are designated as truncated or censored, depending on the nature of the restriction. Samples in which the number of restricted (eliminated) observations is unknown are described as truncated. Those in which restrictions permit counting but not measuring specimens having values outside an interval of measurement are described as censored. Samples of both types frequently occur in life testing, dosage-response determinations, target analyses, biological assays, and in various other investigations.

In the realm of Quality Control, truncated samples are of particular interest in connection with samples from batches or lots of product from which oversized and undersized items have been eliminated as the result of one hundred percentage inspections using go, no-go gages. According to present practice, the effect of this truncation is usually neglected in estimating process (population) means and standard deviations. When gage limits are set at from three to four standard deviations from the process mean, this course of action is justified, but when gage limits are two standard deviations or less from the process mean, neglect of the truncation effect introduces an appreciable bias which causes the process standard deviation to be consistently underestimated.

A large number of research papers and expository articles have been written during the past several years on estimating parameters of various types of populations from truncated and censored samples. For the convenience of readers who desire further information on this subject, some of these are listed as references at the end of this paper, and each of the papers listed contains additional references.

It is the object of the present paper to give a concise account of restricted sampling theory for normally distributed populations and to present techniques for solving estimating equations which apply in the various cases considered. For the benefit of practical minded readers who might be more interested in applications than in theory, several illustrative examples are included. The question of estimate reliability is considered and large sample variances of the estimates are given.

2. SINGLY TRUNCATED SAMPLES

Consider a sample consisting of n observations of a quality characteristic x, (a random variable) such that each observation is subject to the restriction $x \geq x_0$, where x_0 is a known and fixed terminal or truncation point, and the number of otherwise possible observations eliminated as a consequence of this restriction is unknown. When the entire output of some production process is sorted through go, no-go gages and all items for which $x < x_0$ are discarded, random samples subsequently selected from batches or lots of the retained (screened) product are of this type. They are described as singly truncated on the left.

*Prepared in connection with contract DA-01-009-ORD-288, sponsored by the Office of Ordnance Research, U. S. Army.

When x is normally distributed with mean m and standard deviation σ, its frequency (probability density) function is

$$(1) \qquad f(x) = [\sigma\sqrt{2\pi}]^{-1}\exp -(x - m)^2/2\sigma^2, \quad -\infty \leq x \leq \infty,$$

and the likelihood function of a singly truncated random sample from this population when the terminal is to the left, may be expressed as

$$(2) \qquad P = I_0^{-n}[\sigma\sqrt{2\pi}]^{-n}\exp[-\Sigma_1^n(x_i - m)^2/2\sigma^2],$$

where n is the number of sample observations (for which $x \geq x_0$) and I_0 is the proportion of the population from which measured observations are possible.

Let ξ designate the terminal (point of truncation) in standard units of the population,

$$(3) \qquad \xi = (x_0 - m)/\sigma,$$

and I_0 expressed as a function of ξ, becomes

$$(4) \qquad I_0(\xi) = \int_\xi^\infty \phi(t)dt, \quad \text{where} \quad \phi(t) = (\sqrt{2\pi})^{-1}\exp -t^2/2.$$

Taking natural logarithms of (2), and writing L for Ln P, we have

$$(5) \qquad L = -n \ln I_0 - n \ln\sigma - n \ln\sqrt{2\pi} - \Sigma_1^n(x_i - m)^2/2\sigma^2.$$

To derive maximum likelihood estimates (estimators) of m and σ, we differentiate (5) and equate to zero, thereby obtaining

$$(6) \qquad \begin{aligned} \frac{\partial L}{\partial m} &= -\frac{nZ}{\sigma} + \frac{1}{\sigma}\Sigma_1^n\left(\frac{x_i - m}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= -\frac{n\xi Z}{\sigma} - \frac{n}{\sigma} + \frac{1}{\sigma}\Sigma_1^n\left(\frac{x_i - m}{\sigma}\right)^2 = 0, \end{aligned}$$

where Z, the reciprocal of Mill's ratio, is a function of ξ defined as

$$(7) \qquad Z(\xi) = \phi(\xi)/I_0(\xi).$$

From equation (6) it follows that

$$(8) \qquad \begin{aligned} \Sigma_1^n(x_i - m)/n &= \sigma Z, \\ \Sigma_1^n(x_i - m)^2/n &= \sigma^2[1 + \xi Z]. \end{aligned}$$

Let ν_k designate the kth sample moment about x_0; i.e.

$$(9) \qquad \nu_k = \Sigma_1^n(x_i - x_0)^k/n.$$

Using this notation and substituting $m = x_0 - \sigma\xi$, which follows from (3), equations (8) simplify to

$$
(10) \qquad
\begin{aligned}
\nu_1 &= \sigma\,[Z - \xi],\\
\nu_2 &= \sigma^2[1 - \xi(Z - \xi)].
\end{aligned}
$$

Eliminating σ between these two equations gives

$$
(11) \qquad \frac{1 - \xi(Z - \xi)}{(Z - \xi)^2} - \frac{\nu_2}{\nu_1^{\,2}} = 0,
$$

in which ξ is the only independent variable. Accordingly, with the aid of a table of normal curve areas and ordinates, and with $\nu_2/\nu_1^{\,2}$ computed for a given sample, standard iterative procedures can be employed to solve equation (11) for the maximum likelihood estimate, $\hat{\xi}$. From the first equation of (10) it then follows that

$$
(12) \qquad \hat{\sigma} = \nu_1/[\hat{Z} - \hat{\xi}],
$$

and from (3), we obtain

$$
(13) \qquad \hat{m} = x_0 - \hat{\sigma}\hat{\xi}.
$$

The symbol (^) serves to distinguish maximum likelihood estimates from parameters estimated, and $\hat{Z}$ is a shortened notation for $Z(\hat{\xi})$.

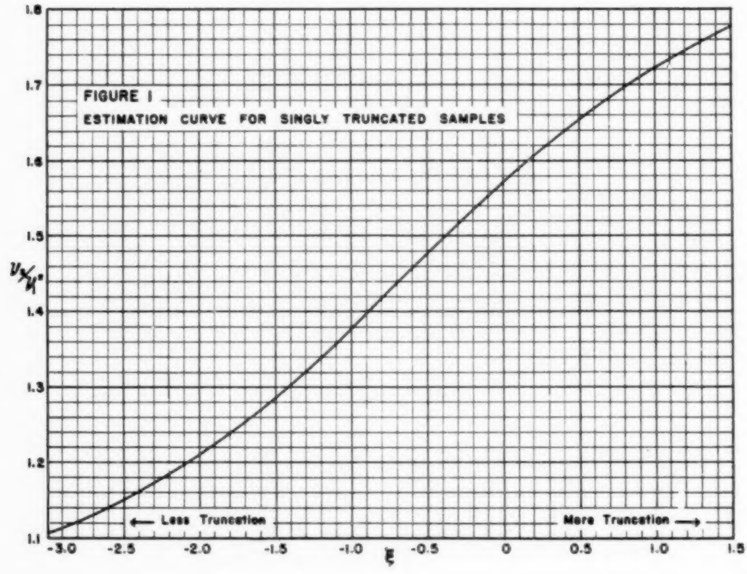
When available, tables of $[1 - \xi(Z - \xi)]/(Z - \xi)^2$ (multiplied by $\frac{1}{2}$) and of $1/(Z - \xi)$ given in reference (4) greatly facilitate the solution of equations (11) and (12). Specimen entries from these tables are

TABLE I. THE FUNCTIONS $1/(Z - \xi)$ AND $m_2/2m_1^{\,2}$

ξ	$1/(Z-\xi)$	$m_2/2m_1^2$	ξ	$1/(Z-\xi)$	$m_2/2m_1^2$	ξ	$1/(Z-\xi)$	$m_2/2m_1^2$
−2.50	0.3971 9772	0.5753 8016	−2.30	0.4294 3629	0.5860 5950	−2.10	0.4662 4750	0.5982 5324
−2.49	.3987 1031	.5758 7929	−2.29	.4311 6321	.5866 3273	−2.09	.4682 1906	.5989 0346
−2.48	.4002 3291	.5763 8201	−2.28	.4329 0163	.5872 0976	−2.08	.4702 0366	.5995 5755
−2.47	.4017 6561	.5768 8833	−2.27	.4346 5162	.5877 9061	−2.07	.4722 0139	.6002 1551
−2.46	.4033 0847	.5773 9828	−2.26	.4364 1328	.5883 7528	−2.06	.4742 1231	.6008 7733
−2.45	0.4048 6157	0.5779 1186	−2.25	0.4381 8666	0.5889 6377	−2.05	0.4762 3651	0.6015 4302
−2.44	.4064 2497	.5784 2909	−2.24	.4399 7185	.5895 5609	−2.04	.4782 7405	.6022 1257
−2.43	.4079 9876	.5789 4998	−2.23	.4417 6893	.5901 5225	−2.03	.4803 2503	.6028 8598
−2.42	.4095 8299	.5794 7454	−2.22	.4435 7797	.5907 5225	−2.02	.4823 8952	.6035 6324
−2.41	.4111 7776	.5800 0278	−2.21	.4453 9905	.5913 5610	−2.01	.4844 6759	.6042 4435
−2.40	0.4127 8313	0.5805 3471	−2.20	0.4472 3224	0.5919 6380	−2.00	0.4865 5932	0.6049 2930
−2.39	.4143 9917	.5810 7035	−2.19	.4490 7762	.5925 7535	−1.99	.4886 6479	.6056 1810
−2.38	.4160 2597	.5816 0970	−2.18	.4509 3528	.5931 9077	−1.98	.4907 8407	.6063 1073
−2.37	.4176 6359	.5821 5278	−2.17	.4528 0528	.5938 1004	−1.97	.4929 1724	.6070 0719
−2.36	.4193 1211	.5826 9960	−2.16	.4546 8771	.5944 3318	−1.96	.4950 6438	.6077 0746
−2.35	0.4209 7160	0.5832 5017	−2.15	0.4565 8266	0.5950 6023	−1.95	0.4972 2557	0.6084 1156
−2.34	.4226 4214	.5838 0449	−2.14	.4584 9014	.5956 9105	−1.94	.4994 0087	.6091 1946
−2.33	.4243 2380	.5843 6257	−2.13	.4604 1030	.5963 2579	−1.93	.5015 9020	.6098 3091
−2.32	.4260 1666	.5849 2443	−2.12	.4623 4320	.5969 6440	−1.92	.5037 9414	.6105 1164
−2.31	.4277 2080	.5854 9007	−2.11	.4642 8890	.5976 0689	−1.91	.5060 1226	.6112 6591

$$m_2/m_1^{\,2} = [1 - \xi(Z - \xi)]/(Z - \xi)^2$$

reproduced here as Table 1. For use as a time saver when accuracy of only one or two decimals is required, a graph of the estimating function of equation (11) is given in Figure 1, below.



For samples that are singly truncated on the right at a point x_0, we need merely recognize that truncation of $f(x)$ on the right at x_0 is equivalent to truncation of $f(-x)$ on the left at $-x_0$. Therefore, in this case we simply change signs of all observations and proceed as when truncation is on the left.

3. DOUBLY TRUNCATED SAMPLES

Doubly truncated samples represent a simple generalization of the singly truncated samples of the preceding section. As the title indicates, a doubly truncated sample is truncated as two points. In this case, let x_0 designate the lower (left) truncation point and w the truncated range. The upper (right) truncation point is accordingly designated as $x_0 + w$. Samples from batches or lots of product screened to eliminate not only items below a fixed lower limit, but also those above a fixed upper limit, are of this type. The logarithm of the likelihood function of a random sample of n fully measured observations from a population distributed according to equation (1) when each observation is subject to the restriction $x_0 \leq x \leq x_0 + w$, and the number of possible observations thus eliminated is unknown, can be expressed as

$$(14)\quad L = -n \ln[I_0(\xi_1) - I_0(\xi_2)] - n \ln\sigma - n \ln\sqrt{2\pi} - \Sigma_1^n (x_i - m)^2/2\sigma^2,$$

where ξ_1 and ξ_2 are left and right terminals respectively expressed

in standard units of the population; i.e.

$$(15) \qquad \xi_1 = (x_0 - m)/\sigma \quad \text{and} \quad \xi_2 = (x_0 + w - m)/\sigma \ .$$

On differentiating (14), equating to zero, and simplifying, we obtain the estimating equations

$$(16) \quad \begin{aligned} &\text{a.} \quad [\bar{Z}_1 - \bar{Z}_2 - \xi_1]/(\xi_2 - \xi_1) \quad - \quad \nu_1/w \quad = 0, \\ &\text{b.} \quad [1 + \xi_1\bar{Z}_1 - \xi_2\bar{Z}_2 - (\bar{Z}_1 - \bar{Z}_2)^2]/(\xi_2 - \xi_1)^2 - \bar{s}^2/w^2 = 0, \end{aligned}$$

where

$$(17) \quad \bar{Z}_1(\xi_1, \xi_2) = \frac{\phi(\xi_1)}{I_0(\xi_1) - I_0(\xi_2)}, \quad \bar{Z}_2(\xi_1, \xi_2) = \frac{\phi(\xi_2)}{I_0(\xi_1) - I_0(\xi_2)},$$

and $\bar{s}^2 = \nu_2 - \nu_1^2$, is the truncated sample variance.

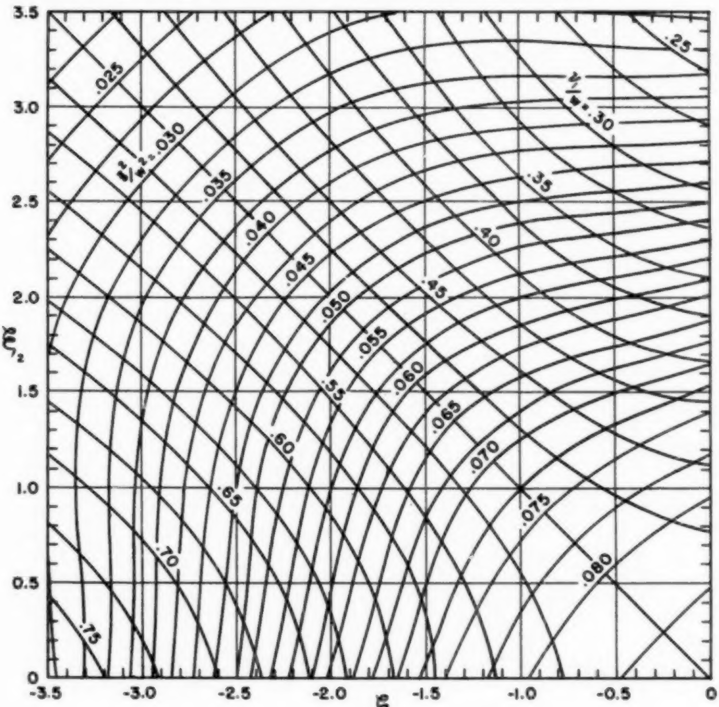


FIGURE 2. ESTIMATION CURVES FOR DOUBLY TRUNCATED SAMPLES

1. Locate curve corresponding to sample value of ν_1/w. If necessary, interpolate. 2. Follow curve thus located to point where it intersects with curve for sample value of $\bar{s}^2/w^2$. Interpolate here also, if necessary. 3. Read the required values of ξ_1 and ξ_2 on scales along the base and left edge of chart as coordinates of the point of intersection determined in (2).

Using standard iterative procedures, the two equations of (16) can be solved simultaneously for estimates $\hat{\xi}_1$ and $\hat{\xi}_2$ with the aid of a table of normal curve areas and ordinates. With these values determined, it follows from equation (15) that

$$(18) \qquad \hat{\sigma} = w/(\hat{\xi}_2 - \hat{\xi}_1), \quad \text{and} \quad \hat{m} = x_0 - \hat{\sigma}\,\hat{\xi}_1.$$

To facilitate solution of (16), the two functions defined therein were tabulated by Thomson (10) for a 0.5 interval of the two arguments. A chart of the two families of curves involved, prepared from his tables, is included in reference (3), and a portion of this chart is reproduced here as Figure 2. Estimates $\hat{\xi}_1$ and $\hat{\xi}_2$ can be read directly from the chart with an accuracy of three to five units in the second decimal. When greater accuracy is necessary, the interpolative procedure illustrated in Section 6 is satisfactory for improving these initial approximations.

The forms in which equations (16) appear here were first suggested by Thomson (loc. cit.). Derivations of an alternate equivalent pair of estimating equations are given in reference (2) in somewhat greater detail than here. We note that the singly truncated estimating equations of the preceding section can also be obtained as a special case of (16) by letting $\xi_2 \to \infty$, since $\lim_{\xi_2 \to \infty} \overline{Z}_1 = Z(\xi_1)$ and $\lim_{\xi_2 \to \infty} \overline{Z}_2 = 0$.

4. CENSORED SAMPLES

As an example of a censored sample, consider a life test which is terminated before all items under test expire, so that of the remaining unexpired specimens, only their number and the fact that their life spans exceed the terminal value is known. Censored samples also arise in connection with dosage-response studies and in various instances, where because of instrument limitations, measurements beyond certain threshold values are not possible although the number of unmeasured specimens can be determined.

With respect to terminal classification, censored samples are of two types, those with fixed terminals and those with variable terminals. The <u>fixed terminal</u> types are the result of sampling complete populations until a fixed number, say n, of the specimens having values within an interval of measurement defined by fixed terminals have been selected and measured. The total number of observations in such samples N, is thus a random variable with possible values n, n+1, n+2, ... In the case of a doubly censored sample with fixed terminals, x_0 and $x_0 + w$, we let n_1 designate the number of unmeasured observations for which it is known only that $x < x_0$, n_2 the number of unmeasured observations for which it is known only that $x > x_0 + w$, and n the number of measured observations for which $x_0 \leq x \leq x_0 + w$. Although n is a fixed number, n_1, n_2, and N ($= n + n_1 + n_2$) are random variables. The logarithm of the likelihood function of a sample of this type from a population distributed according to (1) may be written as

$$(19) \quad L = n_1 \ln[1 - I_0(\xi_1)] + n_2 \ln I_0(\xi_2) - n \ln\sigma - \sum_1^n (x_i - m)^2/2\sigma^2 + \text{const.}$$

Differentiating (19), equating to zero, and simplifying, yields as estimating equations for this case

$$(20)\quad \begin{aligned} &[Y_1 - Y_2 - \xi_1]/(\xi_2 - \xi_1) && - \nu_1/w = 0, \\ &[1 + \xi_1 Y_1 - \xi_2 Y_2 - (Y_1 - Y_2)^2]/(\xi_2 - \xi_1)^2 && - \bar{s}^2/w^2 = 0, \end{aligned}$$

where

$$(21)\qquad Y_1 = [n_1/n]Z(-\xi_1), \text{ and } \quad Y_2 = [n_2/n]Z(\xi_2).$$

Let $\xi_2 \rightarrow \infty$ and the doubly censored sample described above becomes singly censored on the left. After some simplification, estimating equations (20) in this case reduce to

$$(22)\quad \begin{aligned} \nu_1 &= \sigma[Y - \xi], \\ \nu_2 &= \sigma^2[1 - \xi(Y - \xi)], \end{aligned}$$

where the subscript has been dropped from both ξ_1 and Y_1. When σ is eliminated between the two equations of (22), we have

$$(23)\qquad \frac{1 - \xi(Y - \xi)}{(Y - \xi)^2} - \frac{\nu_2}{\nu_1^2} = 0,$$

which can be solved for $\hat{\xi}$ using interpolative procedures similar to those employed in solving (11) in the truncated case. We note that both singly and doubly censored estimating equations are completely analogous with corresponding equations for truncated samples. They differ only in the substitution of Y for Z. Methods suitable for solving truncated sample estimating equations are equally satisfactory for solving censored sample equations. First approximations in the censored cases can be obtained from the curves of Figures 1 and 2, and subsequently improved using interpolative or other iterative procedures. All of the functions involved can be evaluated from tables of normal curve areas and ordinates. In some cases, tables given by Hald (7)* may facilitate the solution of estimating equation (23) for singly censored samples, but they are not essential. These tables give the standardized terminal ξ, (designated as z by Hald) as a function of the double arguments $y = \frac{1}{2}(\nu_2/\nu_1^2)$ and $h = n_1/(n_1 + n)$ for $y = .500(.005)1.500$ and $h = .05(.05).50$.

Samples that are singly censored on the right may be handled in a manner similar to that employed with samples that are singly truncated on the right. When x is replaced with $-x$, we obtain an equivalent transformed sample that is censored on the left.

Variable terminal samples result when both the number of measured observations and the number of unmeasured observations, but not the terminals are fixed in advance of sampling. Gupta(6) pointed out that maximum likelihood estimators for samples of this type from complete normally distributed populations are identical in form with those obtained above for fixed terminal samples when the largest and the smallest of the measured observations are taken as terminals.

5. SAMPLING ERRORS OF ESTIMATES

The variance-covariance matrix of $(\hat{m}, \hat{\sigma})$ is derived from expected values of the second order partial derivatives of L. If we designate

*Also included in Hald's "Statistical Tables and Formulas," John Wiley and Sons, New York (1952).

$-(\sigma^2/n)E(\partial^2L/\partial m^2)$, $-(\sigma^2/n)E(\partial^2L/\partial m\partial\sigma)$, and $-(\sigma^2/n)E(\partial^2L/\partial\sigma^2)$, $\phi_{11}(\xi_1,\xi_2)$, $\phi_{12}(\xi_1,\xi_2)$ and $\phi_{22}(\xi_1,\xi_2)$ respectively, asymptotic (large sample) variances and the coefficient of correlation between estimates may be expressed as

$$
\begin{aligned}
&\mathrm{Var}(\hat{m}) \sim [\hat{\sigma}^2/n][\hat{\phi}_{22}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)],\\
(24)\qquad &\mathrm{Var}(\hat{\sigma}) \sim [\hat{\sigma}^2/n][\hat{\phi}_{11}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)],\\
&\rho_{\hat{m},\hat{\sigma}} \sim -\hat{\phi}_{12}/\sqrt{\hat{\phi}_{11}\hat{\phi}_{22}}\,.
\end{aligned}
$$

Results given here relate to sampling errors of $(\hat{m},\hat{\sigma})$ and they differ accordingly from earlier results of reference (2) which relate to sampling errors of $(\hat{\xi}_1,\hat{\sigma})$. The ϕ_{ij} for the different cases considered in this paper are given below.

<u>For Doubly Truncated Samples</u>

$$
\begin{aligned}
&\phi_{11} = 1 - \bar{Z}_1(\bar{Z}_1 - \xi_1) + \bar{Z}_2(\bar{Z}_2 - \xi_2),\\
(25)\qquad &\phi_{12} = \bar{Z}_1[1 - \xi_1(\bar{Z}_1 - \xi_1)] - \bar{Z}_2[1 - \xi_2(\bar{Z}_2 - \xi_2)],\\
&\phi_{22} = 2 + \xi_1\bar{Z}_1[1 - \xi_1(\bar{Z}_1 - \xi_1)] - \xi_2\bar{Z}_2[1 - \xi_2(\bar{Z}_2 - \xi_2)].
\end{aligned}
$$

<u>For Doubly Censored Samples</u>

$$
\begin{aligned}
&\phi_{11} = 1 + Y_1(Y_1 n/n_1 + \xi_1) + Y_2(Y_2 n/n_2 - \xi_2),\\
(26)\qquad &\phi_{12} = Y_1[1+\xi_1(Y_1 n/n_1+\xi_1)] - Y_2[1-\xi_2(Y_2 n/n_2-\xi_2)],\\
&\phi_{22} = 2+\xi_1 Y_1[1+\xi_1(Y_1 n/n_1+\xi_1)] - \xi_2 Y_2[1-\xi_2(Y_2 n/n_2 - \xi_2)].
\end{aligned}
$$

<u>For Singly (Left) Truncated Samples</u>

$$
\begin{aligned}
&\phi_{11} = 1 - Z(Z - \xi), \qquad \phi_{12} = Z[1 - \xi(Z - \xi)],\\
(27)\qquad &\phi_{22} = 2 + \xi Z[1 - \xi(Z - \xi)].
\end{aligned}
$$

<u>For Singly (Left) Censored Samples</u>

$$
\begin{aligned}
&\phi_{11} = 1 + Y(Yn/n_1 + \xi), \qquad \phi_{12} = Y[1 + \xi(Yn/n_1 + \xi)],\\
(28)\qquad &\phi_{22} = 2 + \xi Y[1 + \xi(Yn/n_1 + \xi)].
\end{aligned}
$$

Although the calculations in some cases are lengthy, the various ϕ_{ij} can be evaluated from tables of normal curve areas and ordinates. When available, Sampford's tables (9) of Z, $Z(Z - \xi)$, and $Z[1 - \xi(Z - \xi)]$ reduce the computing effort otherwise required. Sampford's notation differs from that used here. He writes η for the argument rather than ξ, and he lets $\nu = Z$, $\lambda = Z(Z - \xi)$ with $\zeta = Z[1 - \xi(Z - \xi)]$. In using his tables, however, it is necessary to correct an unfortunate printing error which resulted in negative signs before some of the entries for ζ.

6. ILLUSTRATIVE EXAMPLES

<u>Example No. 1, Sample Singly Truncated on Left.</u> To insure meeting a lower specification of 0.1215 in. on the thickness of a certain insulating washer, the entire production is sorted through go, no-go gages and all non-conforming washers are eliminated. For a random sample of 100 washers selected from the screened production, $\sum_1^n (x_i - x_0) = 0.3124$ and $\sum_1^n (x_i - x_0)^2 = 0.001187$, with $x_0 = 0.1215$. Since $n = 100$, we have $\nu_1 = 0.003124$, $\nu_2 = 0.00001187$, and $\frac{1}{2}(\nu_2/\nu_1^2) = 0.60813314$. By inverse interpolation in Table 1, we obtain $\xi = -1.955$, and by direct interpolation, we find $1/(\hat{Z} - \xi) = 0.495642$. From equation (12) it then follows that $\hat{\sigma} = (0.003124)(0.495642) = 0.00155$, and from equation (13) $\hat{m} = 0.1215 - (-1.955)(0.00155) = 0.1245$. Using equations (24) and (27), we compute $\sigma_{\hat{m}} = \sqrt{V(\hat{m})} \sim 0.000172$, $\sigma_{\hat{\sigma}} = \sqrt{V(\hat{\sigma})} \sim 0.000135$, and $\rho_{\hat{m},\hat{\sigma}} \sim -0.279$.

<u>Example No. 2, Doubly Truncated Sample.</u> To meet stringent specifications on diameter, the entire production of a certain bushing is sorted using go, no-go gages. All of a diameter in excess of 0.6015 in. and all of a diameter less than 0.5985 in. are discarded. For a random sample of 75 bushings selected from the screened production, $\sum_1^n (x_i - x_0) = 0.1237$ and $\sum_1^n (x_i - x_0)^2 = 0.00023186$ where $x_0 = 0.5985$, $w = 0.0030$, and $x_0 + w = 0.6015$. With $n = 75$, $\nu_1 = 0.00164933$, $\nu_2 = .0000309147$, $s^2 = 0.000000371187$, $\nu_1/w = 0.54978$, and $s^2/w^2 = 0.041242$. Interpolating between the curves of Figure 2 with these two latter values, we read $\xi_1 = -2.50$ and $\xi_2 = 2.00$. In many cases, this degree of accuracy might be sufficient. However, the accuracy of these initial results can be improved using the two-way interpolation which is summarized below. *

ξ_2	ξ_1 by Eq.(16a)	ξ_1 by Eq.(16b)	Diff.
1.950	-2.475	-2.558	+0.083
1.998	-2.526	-2.526	0
2.000	-2.528	-2.525	-0.003

Accordingly, as final estimates, we have $\hat{\xi}_1 = -2.526$ and $\hat{\xi}_2 = 1.998$. From equation (18), we compute $\hat{\sigma} = 0.0030/[1.998 - (-2.526)] = 0.000663$ and $\hat{m} = 0.5985 - (0.000663)(-2.526) = 0.6002$. As measures of reliability of these estimates, we employ equations (24) and (25) to compute $\sigma_{\hat{m}} \sim 0.0000840$, $\sigma_{\hat{\sigma}} \sim 0.0000726$, and $\rho_{\hat{m},\hat{\sigma}} \sim +0.151$.

<u>Example No. 3, Sample Singly Censored on Left.</u> A certain breaking strength test is performed by applying an initial (minimum) stress and then gradually increasing it until the test specimen fails. To save time the initial stress x_0 may be established high enough that an occasional specimen fails under this minimum stress. For such censored readings, it is known only that the breaking strength is less than or at most equal to x_0. Individual readings are obtained on all specimens for which $x > x_0$. The resulting samples are thus singly censored on the left at x_0. There is, however, one minor difference between samples involved here and those discussed in Section 4. Here the terminal is included in the interval where censoring occurs. In Section 4, the terminal was included in the interval of measurement. The maximum likelihood estimating equations turn out to be identical in the two cases so the difference is not important. For a hank strength test of the above type performed on a woolen

---

*For further details of this method, c.f. Whittaker and Robinson, "The Calculus of Observations," Blackie and Sons, London (1929), pp. 88-91.

yarn with an initial stress of $x_0 = 70.0$ lbs. being applied, $\Sigma_1^n (x_i - x_o) = 532.7$, $\Sigma_1^n (x_i - x_o)^2 = 7262.13$, the number of measured observations, $n = 50$, and the number of censored observations in which specimens failed on application of the initial stress, $n_1 = 3$. Accordingly, $\nu_1 = 10.654$, $\nu_2 = 145.2426$, and $\nu_2/\nu_1^2 = 1.2795835$. Reading from Figure 1, we have $\xi = -1.55$ as a first approximation. The interpolation involved in subsequently completing the solution of equation (23) is summarized below.

ξ	$[1 - \xi(Y - \xi)]/(Y - \xi)^2$
-1.550	1.2878142
-1.570	1.2795835
-1.600	1.2669025

Thus as final estimates, we have $\hat{\xi} = -1.570$. From the defining relation (21), we evaluate $Y(-1.57) = 0.119905$ and $Y(-1.570) - (-1.570) = 1.689905$. It follows from equation (22) that $\hat{\sigma} = \nu_1/(\hat{Y} - \hat{\xi})$, and thus we compute $\hat{\sigma} = 10.654/1.689905 = 6.304$ lbs. It follows from (13) that $\hat{m} = 70.0 - 6.304(-1.570) = 79.90$ lbs. As measures of estimate reliability, we compute $\sigma_{\hat{m}} \sim 0.870$, $\sigma_{\hat{\sigma}} \sim 0.641$, and $\rho_{\hat{m},\hat{\sigma}} \sim -0.0275$.

REFERENCES

1. COHEN, A. C., JR., "On estimating the mean and standard deviation of truncated normal distributions," J. Am. Stat. Assn., Vol. 44 (1949), pp. 518-25.

2. COHEN, A. C., JR., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," Ann. Math. Stat., Vol. 21 (1950), pp. 557-69.

3. COHEN, A. C., JR., "Truncated and censored samples," Tech. Report No. 11, Contract DA-01-009-ORD-288, Math. Dept., Univ. of Ga.,(1954).

4. COHEN, A. C., JR. and WOODWARD, JOHN, "Tables of Pearson-Lee-Fisher functions of singly truncated normal distributions," Biometrics, Vol. 9 (1953), pp. 489-97.

5. FISHER, R. A., "Properties of H_h functions," Math. Tables, Vol. 1, British Assn. for Advancement of Science (1931), pp. xxvi-xxxv.

6. GUPTA, A. K., "Estimation of the mean and standard deviation of a normal population from a censored sample," Biometrika, Vol. 39 (1952), pp. 260-73.

7. HALD, A., "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," Skandinavisk Aktuarietidskrift, Vol. 32 (1949), pp. 119-34.

8. PEARSON, KARL and LEE, ALICE, "On the generalized probable error in multiple normal correlation," Biometrika, Vol.6 (1908),pp. 59-68.

9. SAMPFORD, M. R., "The estimation of response-time distributions, Part II," Biometrics, Vol. 8 (1952), pp. 307-69.

10. THOMSON, G. W., FRIEDMAN, M., and GARELIS, E., " Estimation of mean and variance of normal populations from doubly truncated samples," Ethyl Corporation Research Report, (1954).

THE PROCESS OF LEARNING BY EXPERIMENT

Eugene W. Pike
Lincoln Laboratory
Massachusetts Institute of Technology

By 1940, the control chart had been perfected by Shewhart, while Dodge and Romig had published the technique of sampling acceptance tests. The history of quality control since that time has been primarily one of great achievement in pioneering the organizational forms and techniques by which the potentialities of these methods could be realized through informed, objective control of manufacturing processes, and in establishing liaison with executive and operating personnel. I think it is safe to say that the recent general acceptance of statistical experiment design, operations research, and many other new management tools could not have come about except for this pioneering work by quality control engineers. On the other hand, this intensive effort has tended to deflect attention away from the theoretical aspects of the subject.

In the last few years there has been a revival of interest in new quality control techniques based on sequential analysis, decision theory, statistical experiment design, and time series analysis. I would like to go along with this interest by talking about two ideas which first appeared during the development of quality control, and which have since had a flourishing growth of their own in other fields. It seems likely that quality control might profitably meet its grown-up children. The first of these ideas is the process by which a scientist (or anyone else) learns from experiments; the second is the recognition of meaningful signals which are almost obscured by noise.

Perhaps the simplest introduction to the first idea would be a brief review of the familiar process of mass manufacture. Figure 1 shows the logical structure of this process in the form of an information flow chart. Mass manufacture begins with a design (the upper left-hand box) which is essentially a statement of what we intend to manufacture. From the design, engineers prepare a <u>manufacturing plan</u>, a detailed set of instructions for realizing the design by manufacture. Such and such raw stock, such and such machines, such and such processing and finishing. Other engineers, at the same time, derive from the design a set of <u>specifications</u> which state what the results of inspection should be, if the manufacturing plan does realize the design.

The next step is the <u>manufacturing operation</u>, which results in a large number of nearly identical pieces of <u>product</u>. (These pieces may be regarded, from a statistical point of view, as random samples from the population which would be made if the manufacturing plan were repeated an infinite number of times. This little fiction has no application here, but will be useful later.) These pieces are <u>inspected</u>, and the resulting measured properties define the actual, as opposed to the intended, manufacturing plan. These properties are <u>compared</u> with the specifications, and the product is either accepted and <u>shipped</u> if the two agree, or rejected and scrapped or reworked if they do not agree. Extensive rejections usually lead to a modification of the manufacturing plan, as well.

In addition, the pieces of product are compared with each other, in the order of manufacture, by means of a control chart, and from the information so gained the manufacturing operation is brought into a state of statistical <u>control</u>.

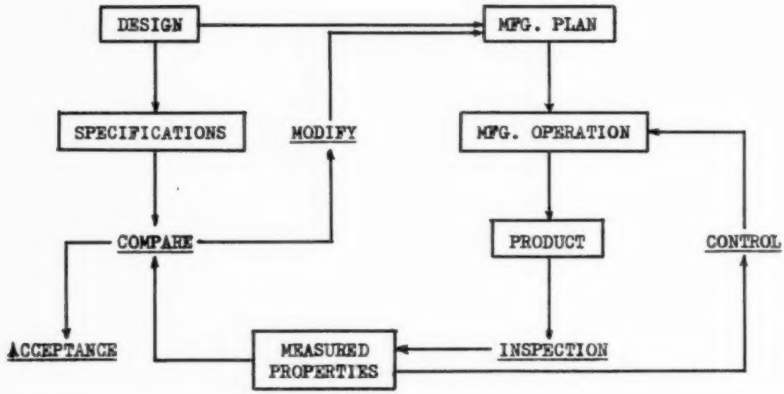


Figure 1.

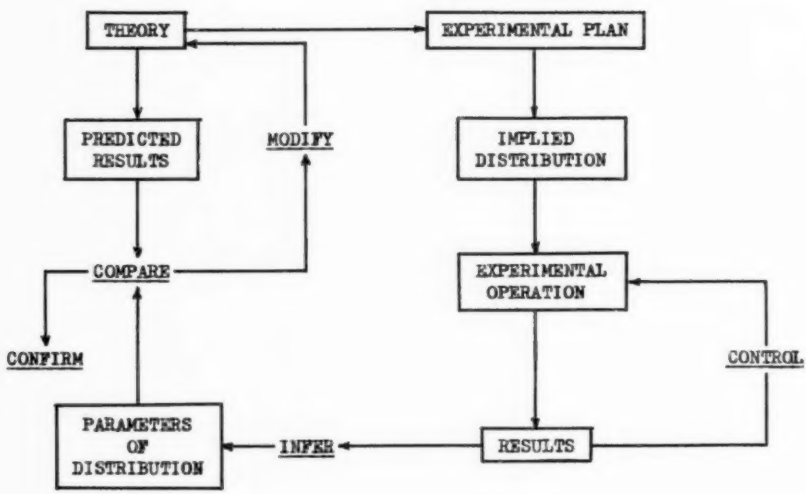


Figure 2.

This flow chart discloses that the manufacturing process can be regarded as a pair of servo loops, in which information about the result of the manufacturing operation is returned to the operation, and to the manufacturing plan as well, so that the resulting product conforms to the specifications. Now these servo loops have the same inherent possibilities of instability and oscillation, or superstability on the other hand, which any similar loops in feedback amplifiers or process control systems have. Up to now, the adjustment of the manufacturing process loop for optimum stability and rate of convergence on the design has been on a cut-and try-basis, but I feel that a basic analysis using servo theory and operations research techniques is well within present capabilities, and that it might produce very valuable results. It would at least display the relative functions of standard control chart techniques, sequential techniques, and so forth much more clearly than is presently possible.

Figure 2 is a similar flow chart showing the elementary logical structure of the process of learning by experiment, based on the insights of Galileo, Shewhart, Bridgman, Fisher, and many other scientists. It is plain that it is very similar to the process of mass manufacture, as Shewhart first pointed out in 1939 (1). In the place of the design, the experimental process starts with the <u>theory</u>, which is a model of some aspect of reality. The objective of the process is not to make the results of experiment conform to the theory as a product should conform to the design, but to adjust the theory so that its predications conform to the results of experiment.

To realize this objective, the scientist chooses some significant experiment, and prepares (consciously or subconsciously) a detailed <u>experimental plan</u>, a statement of the precise sequence of operations which defines the experiment. Such and such equipment, arranged so and so, with this and that recorded precisely. At the same time, a set of <u>predicted results</u> are derived from the theory, in direct analogy to the specifications derived from a design, for comparison with the actual results.

An experimental plan can be thought of as implying the results of carrying it out an unlimited number of times. It was first pointed out by Dr. Shewhart that, just as no manufacturing process makes absolutely identical products, the result of attempting to repeat an experimental plan indefinitely must be a distributed population of results covering a finite range of values. Among the causes for this irreducible variation from one experiment to the next are the fact that the experiment must be performed by a finite human being in a finite time, the fact of thermal agitation at any finite temperature, the finite electronic charge and other quantum limitations, the impossibility of finding identical living organisms in biological work, and so on. Usually the result predicted by theory is simply the mean of this <u>implied distribution</u>, although sometimes the entire distribution can be predicted.

The results of actual experimentation are presumably random samples from this implied distribution. From these samples one must <u>infer</u>, by statistical reasoning, the mean of the implied distribution, or whatever <u>parameters of the distribution</u> are to be <u>compared</u> with the predicted results from the theory. If the inferred parameters of the implied distribution agree with the predicted values, the theory is confirmed, and may be used with increased confidence as the basis for new advances, or of engineering designs. If, as is more usual, the two disagree, then it is

necessary to modify the theory until its predictions agree with the results of experiment.

Many of you have surely picked up the "presumably random samples from this implied distribution." One must confess that scientists in laboratories are no more successful in keeping their experiments free from assignable causes of variation, in the past, than production engineers have been with their manufacturing processes. The advantages of better equipment and laboratory conditions are at least balanced by the more difficult operations which the scientist undertakes, and by the fact that large numbers of repetitions are expensive and time consuming. The ordinary procedures of quality control cannot be used unchanged to bring laboratory experiments under control, but Dr. Olmstead and others have shown that statistical quality control methods can be modified to suit the special conditions. There should be an interesting and expandable secondary field for quality control in the laboratory, although both parties will have a great deal to learn about each other before smooth cooperation can be expected.

A very important difference between a manufacturing process and the process of learning by experiment is that the latter is enormously more variable in form. In manufacturing, one always starts with a design, and proceeds to a product which is inspected. The scientists must often start with a series of uncontrolled observations from an inaccessible experimental operation (the astronomers, the biologists, and the social scientists, for example), and construct both a theory and an experimental plan to complete the loop. Again, it is inconceivable that an automobile factory should produce sewing machines, say, against the intent of its managers, whereas systematic errors of this magnitude are common in many of the more difficult fields of science, in spite of the skill and care of the scientist performing the experiment. The result of an experimental plan may be a number, a functional form, a multiple comparison, a discrimination, or some other even more complex pattern.

Because of this enormous variability in form as well as content, the art of making this servo loop converge rapidly and stably, so that theory is enlarged and made more like reality, remains an art. The technique of making the statistical inference from the results of experiment to the comparison with prediction can be systematized, and statistical control methods can help in getting the experimental operation running smoothly, The choice of a crucial experiment, and the construction of fruitful theories, are still attributes of genius which can be taught, if they can be taught at all, only by daily contact between gifted teacher and gifted student.

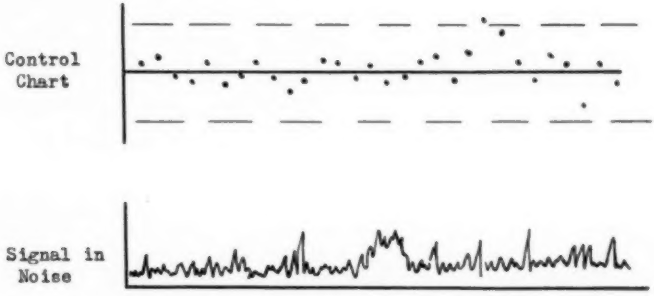
Let us now turn to the second idea. Anyone who is familiar with both communication engineering and quality control methods will be struck immediately by the similarity between a control chart and the oscillograph of a radio signal almost lost in noise. This is shown in Figure 3. On examination, this similarity is more than skin deep. The noise in the electrical circuit is a random, bounded variation which results from the combination of the irregular motions of a great number of electrons under the driving force of thermal agitation. The irregular variation of the points on the control chart is a random, bounded variation resulting from the combination of many small irregular effects in the manufacturing process. The pattern of points on a stable control chart is just as much a "noise", in the strict meaning of the term in

communications engineering, as is the pattern of voltage variation with time across a hot resistor.

The signal in the communications channel is an isolated disturbance of agreed or at least recognizable form, having a single cause. An assignable cause in a manufacturing process produces an isolated disturbance of recognizable form in the control chart. In each case, the objective is to recognize when a signal (or an assignable cause) is present, even when it is small, and difficult to distinguish from the accidental shapes produced by the noise. In each case, mathematical operations are performed on the combination of noise plus possible signal, in order to make recognition easier. In the case of the control chart, the mathematical operations are performed numerically, and plotted on the chart. In communication practice, the mathematical operations are carried out by electrical circuits, called <u>filters</u>. The function of the two is identical.

The recognition of signals obscured by noise is so fundamental to communication practice that there has been a very great development of theory in this field, reaching to the point of basic contributions to the mathematical discipline of decision theory. This is probably not of general or continuous interest to quality control engineers, since the control chart is still a very good filter indeed, but for special situations where the cost of inspection is very high, for instance, one might make good use of the very highly developed methods of the communications engineer.

The literature of this work is naturally in communications terminology, and it takes some effort to translate it into quality control language. If one remembers that "noise" is parallel to the random, bounded variations found on the control chart of a stable process, and that the step or ramp produced by an assignable cause is a signal of that shape, then time and patience will suffice. Middleton and Van Meter (2) provide the most complete and general formulation of the problem, in terms of decision theory, and include an excellent introduction and bibliography. The earlier book of Lawson and Uhlenbeck (3) is less condensed and starts at a more elementary level, but much of the discussion is in terms of specific communications equipment. Finally, Marcum (4) has condensed into graphical form a very comprehensive exploration of the possible combinations of a number of observations, ratio of signal amplitude to noise amplitude, and "false alarm rate." This last is the probability that the system will state that an assignable cause is present when in fact there is only noise.



<u>Figure 3.</u>

Bibliography

1. Shewhart, W. A., "Statistical Theory", U. S. Department of Agriculture Graduate School, Washington 25, D.C., 1939.

2. Van Meter, David and Middleton, David, "Modern Statistical Approaches to Reception in Communication Theory.", Transactions of the Institute of Radio Engineers, PGIT-4, p.119, 1954.

3. Lawson, J. W. and Uhlenbeck, G. E., "Threshold Signals", McGraw-Hill, N. Y., 1951 (Volume 24 of the M.I.T. Radiation Laboratory Series).

4. Marcum, J.I., "A Statistical Theory of Target Detection by Pulsed Radar", RAND Corporation, Memorandum RM-754, 1952 (Reissued).

There is no published presentation of the process of learning by experiment, to the best of the writer's knowledge, in the form given above, although the concept has been circulating generally for some years. How one actually operates this process is the subject of a huge, diffuse, and often contradictory literature. Those wishing to read further in the field are commended to Reference 1 above, and to the following books which offer at least a start and a reasonable survey of the field.

5. Wilson, E. Bright, "An Introduction to Scientific Research, " McGraw-Hill, New York, 1952.

6. Churchman, C. West, "Theory of Experimental Inference," McMillan, 1948.

7. Hadamard, Jacques, "An Essay on the Psychology of Invention in the Mathematical Field," Princeton University Press, 1945.

8. Bridgman, P.W., "The Nature of Physical Theory," Princeton University Press, 1936.

ON THE JOB CONTROLS IN STEEL PIPE MILLS

James A. Curry
Kaiser Steel Corporation

I should like to consider with you the statistical process control program in two different types of pipe mills. A description of these mills and the processes we utilize to produce tubular products will be helpful in understanding the illustrations used.

Steel tubular products are those cylindrical forms designated as pipe or tubing which are generally used for conveying gases or liquids and for a diversity of mechanical and structural purposes. The general terms pipe, tubes and tubing are not sharply defined within the industry and are therefore used interchangeably.

At Fontana, California, Kaiser Steel Corporation is engaged in a fully integrated pipe production program involving both continuous welded steel pipe and electric resistance welded steel pipe.

Electric resistance welded pipe is rolled in sizes from 5 9/16" O.D. to 14" O.D., inclusive on a Yoder type welding unit of latest design. This pipe can be produced in wall thicknesses from .188" to .400" inclusive. The maximum lengths produced are 55 feet.

Continuous welded steel pipe is rolled in nominal sizes from 1/2" to 4" inclusive on a modern continuous weld type mill. It is supplied in 21-foot uniform lengths and in random lengths, either plain end or threaded and coupled and either black or galvanized. Both standard and extra strong weights are produced.

ELECTRIC RESISTANCE WELDING PROCESS

Kaiser Electric Resistance Welded Pipe is produced from cold, flat skelp.

The skelp is first passed through a roller leveler to achieve a smooth, flat surface. From the leveler operation, the skelp undergoes edge cleaning which prepares the metal for good contact with the welding electrodes and insures free passage of the welding current. A thorough cleaning is accomplished by a steel shot blasting process under high pressure.

A perfectly straight welding surface is essential and a uniform width must be maintained throughout the full length of the skelp. To insure this, the skelp is passed through rotary shears which trim both edges to close tolerances immediately before the forming and welding operations.

The skelp is passed from the edge trimmer directly into a series of forming rolls which progressively form it into an open tube. The tube is moved into the welding unit where revolving circular electrodes contact the steel close to each edge and transmit the current which generates the welding heat. By careful control of current, speed and pressure, the edges are bonded to produce a weld of the same strength and properties of the parent metal, extruding just enough metal both inside and outside of the tube to insure a complete weld. The extruded flash is immediately removed by stationary cutters, leaving a smooth wall.

The welded pipe is passed through several stands of rolls which slightly reduce the diameter and insure correct size and straightness. Final roll straightening is done prior to a thorough visual inspection of each length of pipe for surface imperfections. The pipe is then magnetically inspected for weld quality. Following inspection and crush testing, the pipe ends are cut and beveled. While under hydrostatic pressure, the pipe is struck with pneumatic hammers and again checked for possible defects.

CONTINUOUS WELDING PROCESS

In order to produce pipe by the continuous weld process, the steel is rolled in coils containing 185 to 550 feet of skelp depending on the size of the pipe being made. As these coils are paid out one at a time, the skelp passes through a roller leveler which flattens it. When the tail of one coil reaches the flash welding machine, the starting end of the next coil is electric resistance welded to it, thereby forming a continuous ribbon. The skelp is drawn through the gas fired reheating furnace which raises it to a welding temperature in the minimum of thirty seconds. As it leaves the furnace, jets of air impinge on the edges of the skelp, increasing the temperature 100 to 200 degrees up to the mean welding temperature. The skelp then passes through a forming roll. Welding and sizing is completed by ten pairs of grooved rolls arranged in five sets, each set consisting of a pair of vertical and a pair of horizontal rolls.

After the pipe is rolled into shape, it is cut to lengths of approximately 21 feet by means of a flying hot saw. The pipe is then passed through a sizing mill where the final sizing is done and scale is loosened and removed both internally and externally. After final cooling, the pipe goes into the finishing department where it is straightened and the ends finished, followed by hydrostatic testing to specification.

STATISTICAL CONTROLS IN THE ELECTRIC WELD MILL

It should be noted here that percentages used in this paper are not to be taken as reflecting favorably or unfavorably on mill operators in various steel plants. The variables that bring about higher or lower percentages of defective product occurring in the flow of operations in diverse plants constitute individual challenges to operators and engineers in Statistical Quality Control, at each plant. As engineers in Statistical Quality Control, our professional interest lies in finding methods to control defects occurring in the flow of operations rather than in the defects themselves.

We started to investigate the possibilities of establishing a Statistical Quality Control program in the Electric Weld Mill in May, 1952. Preliminary investigation revealed that almost half of all downgraded pipe was downgraded because of inability to meet the crush test requirements. Specifications require that both ends of each piece be crush tested to withstand deflection at the weld of at least one-third of the pipe diameter. It seemed logical then that this was a good place to start. A series of process analysis studies were made to determine the normal expected limits for crush deflection and $\bar{X}$ and R charts were installed for this operation in August, 1952. The purpose of these control charts is to indicate to the Head Welder and to Mill supervision when the process goes out of control and remedial action is indicated.

By close study of "out-of-control" crush test production, some of the major causes of weld failure have been determined. As an example, some out-of-control production was found to be associated with variation in the depth of cut of the inside welding flash. We found that if the inside flash cutter cut too deep, the ability of the weld to withstand crush deflection decreased in a straight line relationship. A regrestion analysis showed that crush deflection decreased .55% for each .001 inch of undercutting.

Process studies of the flash cutting operation showed that there was a normal variation in this operation of .045". Specifications require that not more than .032" of inside flash be retained. So in order to meet this specification, it was mill practice to undercut somewhat. Engineering changes on the flash cutter were made which reduced the inherent variation to .020".

Because it was important that flash cutting be controlled as closely as possible because of both the effect on weld crush and to avoid reworking because of excess flash, an average and range chart was started on this operation. Sub-groups of four consecutive pieces are "miked" at the weld and just outside the cutting areas. Plus or minus differences are charted and adjustments made to the cutting tool as required. As a result, "out-of-control" points on the weld deflection charts associated with flash cutting have practically been eliminated. A comparison of two runs of pipe, one in 1953 and one in 1955 as shown in Fig. 1, indicate the improvement which has been achieved in this operation.

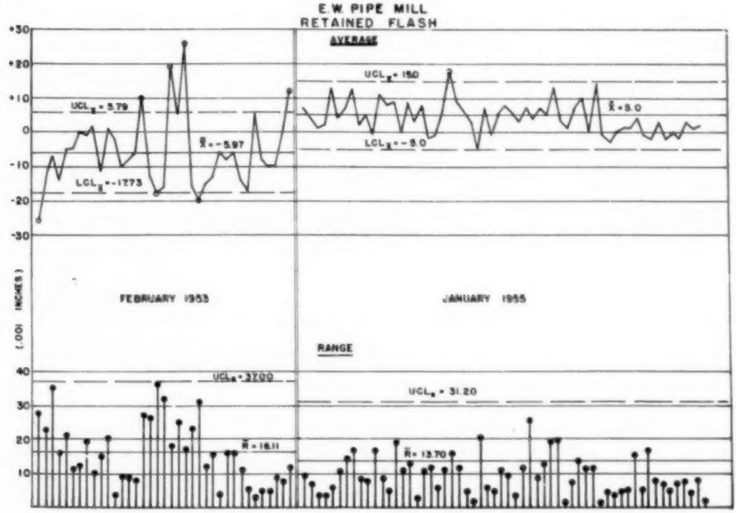


Figure 1

Comparison of Flash Cutting Performance Before and After Establishing Chart Control

Another source of "out-of-control" crush test deflection has been located in variation of sheared width of skelp before forming. A process study covering several eight-hour turns on the rotary side shear showed very close control during much of the time but rather wide fluctuations and changes in average width at other times. Since we still have been unable to eliminate the causes for this shear going out of adjustment, we are sampling the results on $\overline{X}$ and R charts.

Work is continuing to isolate causes of "out-of-control" crush deflection. We now have collected enough evidence to indicate the biggest remaining cause of "out-of-control" crush deflection is associated with variations in shape of the tube at time of welding and with welding pressures. Work is now under way to install strain gage equipment to measure pressures on the vertical welding rolls. This equipment has already been installed to measure electrode pressure. No satisfactory method of measuring welding shape of the tube has yet been developed, so that measurements can be taken on a production basis.

In January, 1952, Kaiser Steel Corporation management decided that for product control purposes, each piece of electric weld pipe would be subjected to magnetic inspection for soundness of the weld zone. It was also determined that no piece of pipe would be classified as prime product when magnaflux indications were present. Most of these magnaflux indications can be ground out and rewelded. However, this is an expensive operation and efficiency of production requires that this defect be held to a minimum.

At the time this problem was first studied, not much factual information was available as to the causes of these weld zone voids. There were as many opinions as to causes as there were people involved. It was recognized that this was not a controlled process and the occurrence of this defect was not normally distributed. For this reason, the idea of P chart control was discarded. Use of C charts based on Poisson was considered and discarded because of the difficulty of counting individual defects in a length of pipe.

A cumulative defects chart was devised with significant changes in rate of production of magnaflux defects being determined as a function of the distribution of the incomplete Beta function at 95% confidence limits. A log of mill operations and changes is also included as a part of this chart. This chart is illustrated in Fig. 2 to indicate the type of information recorded.

Mill operating personnel have found that by changes in mill set up, significant changes in the production of magnaflux indications result. As an example of this, comparing significant changes and electrode trimming, it was found that electrode pressure had a significant effect. Trimming the electrode requires stopping the mill, raising the electrode, trimming, and then lowering the electrode. When this association was discovered, it was determined that some method of adjusting the electrode to the same pressure was required. Strain gage equipment was installed and this source of variation has been controlled.

Other causes of magnaflux indications, but not all of them, have also been isolated. Work is still continuing on this project. Magnaflux indications are currently running well under 10% maximum on any run.

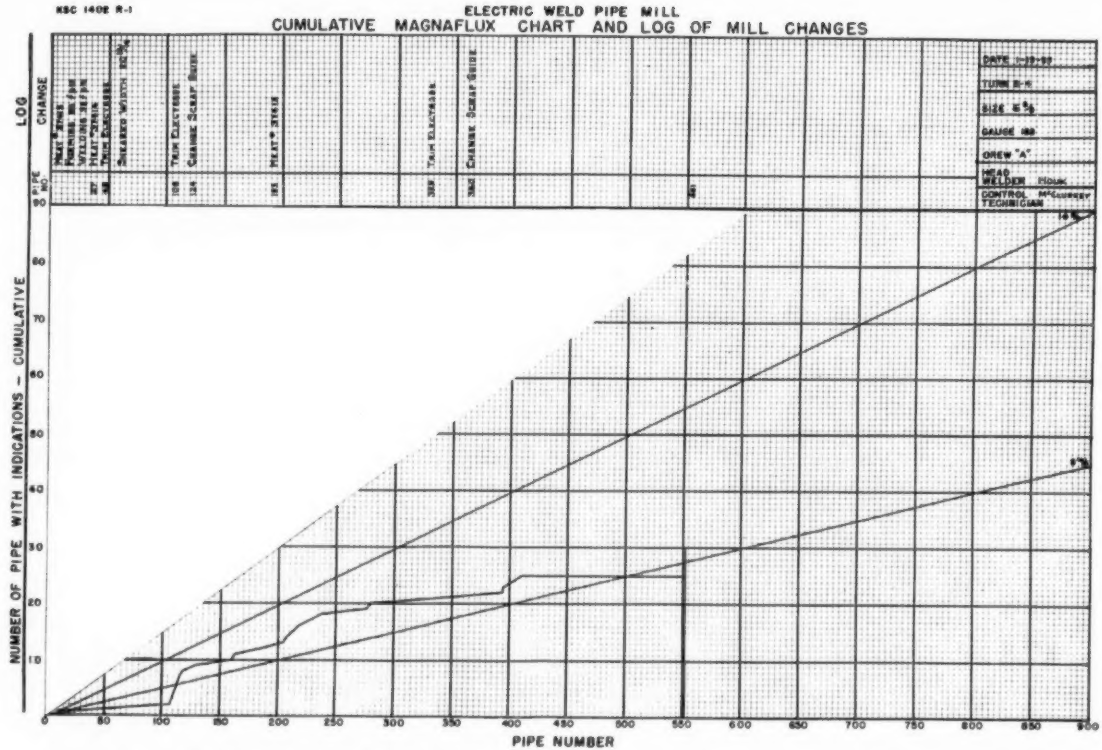


Figure 2 - Cumulative Defects Chart to Determine Significant Changes in Rate of Production of Magnaflux Indications

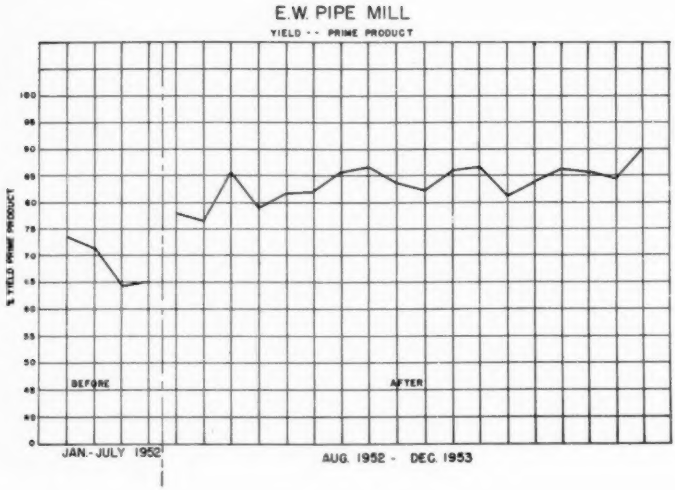


Figure 3

Yield of Prime Product Before and After Installation of S.Q.C. Methods

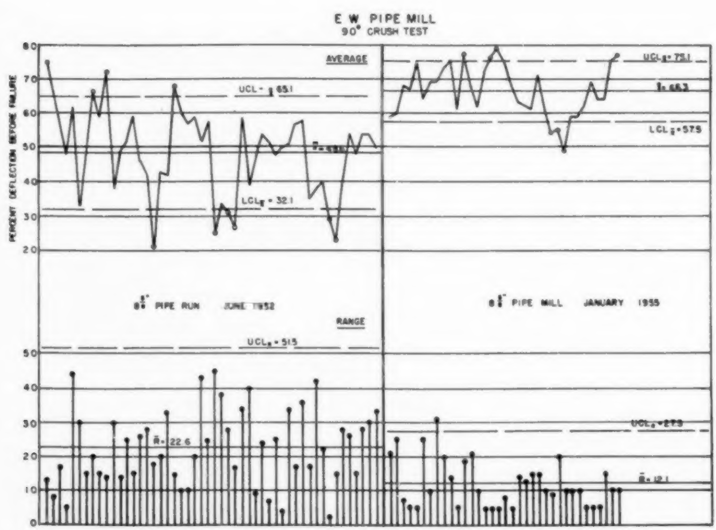


Figure 4

Comparison of Weld Deflection Characteristics of Two Runs of 8 5/8" Pipe - 1952 and 1955

Fig. 3 indicates the extent of the increase which has been made in production of prime product since the quality control program was started in this mill. Fig. 4 illustrates changes since 1952 in average and variability of crush test deflection. The wholehearted cooperation of mill superintendents and supervision was a prime factor in the attainment of these results. We are well aware that S.Q.C. methods in and of themselves are not responsible for any improvements. However, in this case, the S.Q.C. program, coupled with a new emphasis on quality awareness, full cooperation of the Metallurgical Department, along with steady and continuing pressure applied by the mill Superintendent, all happening at the same time, has resulted in a very satisfying improvement in yield and quality.

STATISTICAL CONTROLS IN THE CONTINUOUS WELD MILL

The statistical quality control program in our Continuous Weld Pipe Mill is not as old nor as far advanced as in the Electric Weld Mill. However, by using statistical methods as a means of process control, significant gains have been made in a relatively short time. It would appear from our experience that the application of statistical techniques as a method of controlling both manufacturing and finishing operations in a continuous weld pipe mill more closely approximates conditions found in mass production manufacturing than perhaps any other steel mill operation.

In the manufacture of steel pipe, specifications to which it is manufactured and sold specifically state that each piece will be accepted or rejected on its own merits. For this reason, our statistical efforts have all been slanted toward the objective of control of process level rather than for use as inspection sampling plans.

At the present time, we are using statistical charts as the basis for control of the following characteristics:

Average outside diameter
Ovality or out-of-roundness
Forming defects
Welding defects
Straightener performance
Threader adjustment

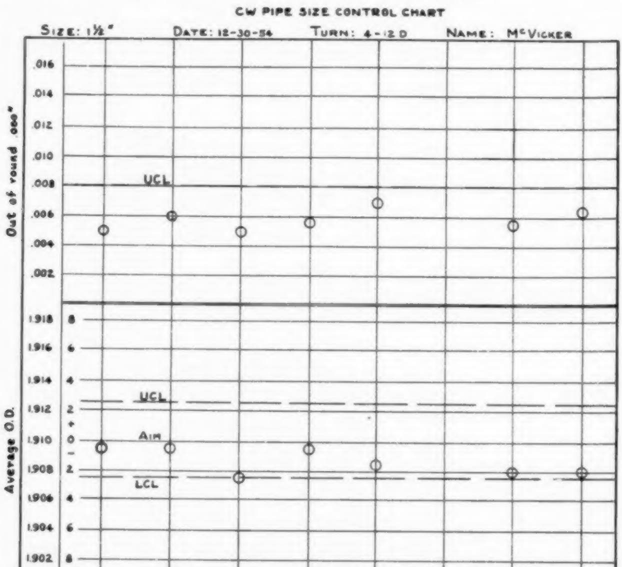
I should like to discuss briefly the objectives of each of these controls, the type of control used and the results obtained.

Average Outside Diameter and Average Out-of-Roundness

After the skelp is heated, formed, welded and cut to length by a flying hot saw, the pipe is sized. It passes over a cooling rack, at the end of which a cold saw cuts the pipe to exact lengths. At this point, a mill inspector makes periodic tests of the weld by flattening. This inspector also makes periodic tests of four successive samples for both average outside diameter and out-of-roundness. The average of the maximum and minimum measurements is plotted as the average O.D. and the average difference in the two measurements is plotted as the average out-of-roundness. These variables are both plotted on regular $\overline{X}$ and R charts with the exception that the out-of-roundness chart has no lower control limit. A typical turn's operation of the out-of-roundness and average O.D. chart is illustrated in Fig. 5. Here is an example of a

case where we have no difficulty at all in staying within the specification tolerances. However, we have found that wide fluctuations in outside diameter, even within specification limits, has a marked effect on the performance of the straighteners. If a straightener operator sets up his machine on the tight side we have a high incidence of damaged ends and spiral rings. If the straightener is set too loosely, then the pipe is not straightened the first time through and has to be re-run. Fig. 6, representing eighty hours of continuous production of 3/4" pipe is shown as an example of the fluctuation in outside diameter which was encountered. All of this production was well within specified tolerances. However, straightener rejects on this run were much higher than expected.

Out-of-roundness or ovality in excess of .010" to .012" has a marked effect on the production of flat threads in the threading operation. For this reason, we hope eventually to control this characteristic to .008". Figures 7 and 8 indicate that definite progress is being made in the closer control of both O.D. and out-of-roundness.



TIME	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
MAX. 1	+2	+3	+1	+3	+2	Mill	0	+3
2	+2	+1	+1	0	+5	Down	+1	+1
3	0	0	0	+2	+1		0	0
4	+2	+3	0	+3	0		+1	0
TOTAL	+6	+7	0	+8	+8		+2	+4
AVE.	+1½	+2	0	+2	+2		+½	+1
Out of round	5	6	5	5½	7		5½	6½
MIN. 1	-2	-2	-5	-2	-7		-5	-2
2	-5	-5	-2	-5	-5		-3	-5
3	-5	-5	-5	-4	-4		-8	-8
4	-2	-4	-7	-3	-3		-5	-7
TOTAL	-14	-16	-19	-14	-19		-21	-22
AVE.	-3½	-4	-5	-3½	-5		-5	-5½
Aver. O.D.	-1	-1	-2½	-½	-1½		-2	-2

Figure 5 - Combined Data Collection Sheet and Control Chart for Average O.D. and Out-of-Roundness

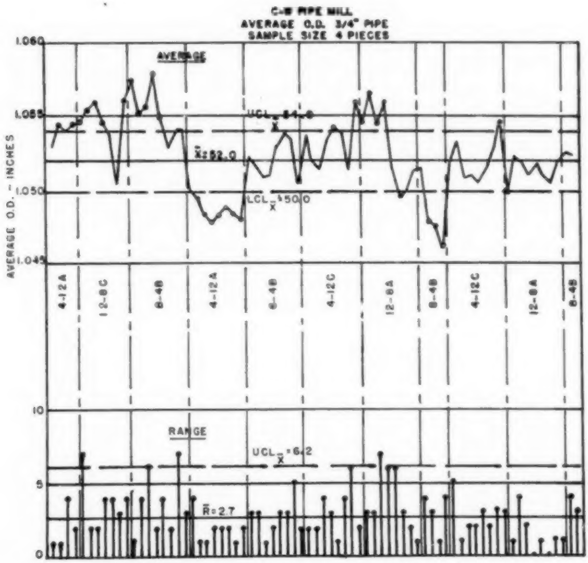


Figure 6 - Eighty Consecutive Hours Operation Showing Typical Fluctuation in O.D. Which Was Encountered Prior to Control. All Production Was Well Inside Specification Limits of 1.018" Minimum and 1.065" Maximum.

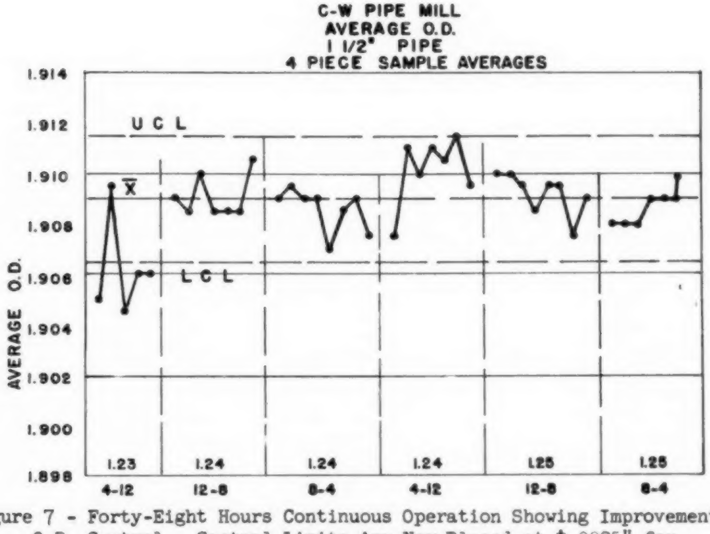


Figure 7 - Forty-Eight Hours Continuous Operation Showing Improvement in O.D. Control. Control Limits Are Now Placed at ±.0025" for Sub-Group Averages of Four Samples.

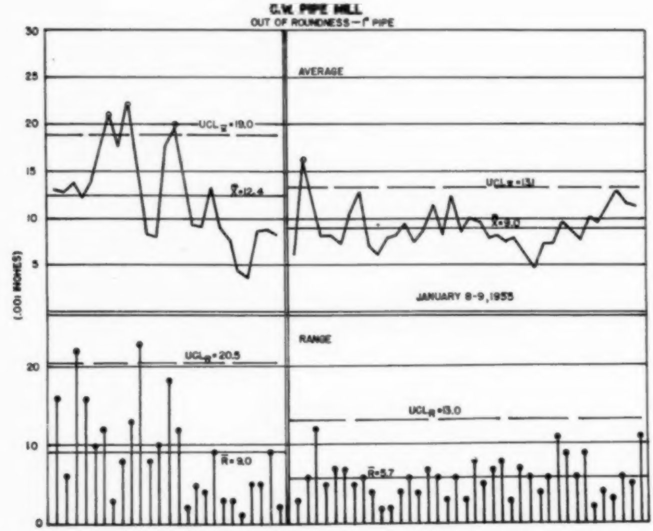


Figure 8 - Out-of-Roundness Has Been Materially Reduced Since Inauguration of Control Methods.

Forming, Welding and Straightener Defects

In order to explain the statistical approach which we have taken on forming and welding defects and on straightener performance, an explanation of our inspection procedure is necessary.

Our former practice was to give all pipe an in-process inspection at a separation bench, after the straightening operation. Manufacture to this point is a "straight line" operation. The purpose of this inspection was to cull out the more obvious defects after straightening so that further processing could be avoided on defective or downgraded pipe, and to present a minimum number of defective pieces to the final inspection operation.

This method of separation bench inspection of all pipe, although it did cull out most of the defective pipe and pipe which needed to be recut and/or restraightened, has many disadvantages. Among these disadvantages four of the most important were that it was a bottle-neck operation which impeded the steady flow of pipe to the finishing operations; that it was expensive; that usually this inspection lagged from three to six turns behind the mill operation which made it impossible to use inspection information gained to control either mill or straightener operation; and finally, due to the good quality of many runs of pipe -- this inspection accomplished very little because of the small fraction defective. It should be understood that this separation bench inspection was in addition to 100 percent final inspection.

The inspection department decided that final inspection would be able to maintain a satisfactory outgoing quality level with an average

of 1-1/2% forming defects, 1-1/2% welding defects and 1-1/2% crooked pipe coming to them.

It was therefore decided that a sampling plan would be developed which would allow production from the mill and from the straighteners to by-pass the separation bench when the average of each of these classifications of defects was 1-1/2%.

This is accomplished by taking periodic samples from current production, inspecting the sample immediately and basing the decision to by-pass the separation bench or to go over the separation bench, depending on the results of the sample. Some compromise was required here in that the sample size of fifty pieces seemed to be the largest practical number which could be handled with assurance that inspection would be on a current basis. For this reason, it is recognized that this sampling plan (N=50 rejection number = 4) does not offer the same degree of assurance on all sizes because of varying production rates. However, experience has shown that the plan does work rather well, even though the operating characteristic curve of this inspection plan is somewhat flatter than we would like it to be.

As each sample is taken and inspected, the inspector informs the straightener operator and the mill foreman of the results of the inspection. In this way we are able to keep the operating units informed as to the current quality level being produced. A sample of the form on which this data is collected for each operating turn is illustrated in Fig. 9.

C.W. PIPE MILL SAMPLING INSPECTION
HOT MILL AND STRAIGHTENERS

DATE 1-20-55 NAME C. Link

TURN 12-8A

								FORMING DEFECTS							WELDING DEFECTS					STRAIGHTENER DEFECTS	HOOKS							
MILL OR HAND FED	STR. UNIT	HEAT NUMBER	TIME SAMPLE TAKEN	WELDED CREW TURN DATE	SIZE AND LENGTH	SAMPLE SIZE	LIFT NUMBER	OFFSET SEAMS	DEEP SEAMS	ROLL MARKS	CAVES			TOTAL	SPLIT ENDS	NOT WELDED			TOTAL	ENTRY DAMAGE	BOTH ENDS	NORTH ENDS	SOUTH END	CROOKED	SPIRAL RINGS	EXIT DAMAGE		TOTAL
MF	NKR	37379	12:30 am	1-18 A 12-8	1¼ x UL	50	63177																	2				2
MF	SKR	37379	1:10 am	1-18 A 12-8	1¼ x UL	50	57389	Adjusted guides and rolls, machine O.K. 1:15AM																		31		31
MF	NKR	37381	2:35 am	1-17 A 12-8	1¼ x UL	50	63221				5			5								2						2
MF	SKR	37381	3:30 am	1-17 A 12-8	1¼ x UL	50	78321				1			1								1						1
MF	NKR	37381	4:40 am	1-17 A 12-8	1¼ x UL	50	63108				1			1									1	1				2
MF	SKR	37381	5:45 am	1-17 A 12-8	1¼ x UL	50	78324				2			2								1						1
MF	NKR	?	6:50 am	?	1¼ x UL	50	78411				2			2										2				2

SUTTON IDLE

SKR						NKR					
	1	63177	1¼"	17	63309 1¼"		1	63177	1¼"	17	[illegible] 1¼"
	2	[illegible]				*	2	[illegible]		18	[illegible]
*	3	[illegible]				*	3	[illegible]	*	19	[illegible]
*	4	[illegible]				*	4	[illegible]	*	20	[illegible]
*	5	[illegible]				*	5	[illegible]		21	
	6	[illegible]	Caves				6	[illegible]	Caves		
*	7	[illegible]				*	7	[illegible]			
*	8	[illegible]				*	8	[illegible]			
	9	[illegible]	Abnormal				9	[illegible]	Caves		
*	10	[illegible]				*	10	[illegible]			
	11	[illegible]				*	11	[illegible]			
*	12	[illegible]					12	[illegible]	Caves		
*	13	[illegible]					13	[illegible]	Caves		
*	14	[illegible]					14	[illegible]	Caves		
	15	[illegible]	Caves				15	[illegible]	Caves		
*	16	[illegible]					16	[illegible]	Caves		

37 lifts straightened
20 lifts bypassed

Figure 9 - Data Collection Form for Determining Forming, Welding and Straightener Defects

The results of this sampling inspection have been very good to date. For example, process studies showed that the amount of pipe that had to be re-run, i.e., restraightened, in order to pass our straightness standards was as high as twenty times the process average under controlled conditions. On a recent run of one size pipe, straightener re-runs were reduced 90% as compared to the last previous run of this same size prior to inauguration of the sampling plan and current feed back of information to straightener operators and mill supervision. It should be pointed out too, that operating supervision at the same time instituted a training program on straightener operation and set-up. Here again, full credit for this exceptional showing must be given to operating personnel because they are the only group that can actually accomplish the changes which will reduce reject or re-process percentages.

Pipe Threader Controls

Pipe threading has always been an operation which produced a high percentage of rejects in our pipe finishing operation. This not only causes a large amount of reprocessing (cutting and rethreading) but also results in a large percentage of random length pipe. Finished yield is also adversely affected.

Machine capability studies involving experimental runs in factorial design convinced us that approximately half of this high reject percentage was associated with variation in the threading equipment and approximately half was associated with the condition of pipe coming to the threaders.

Of all the reasons for threader rejects, flat threads accounted for more than half of the total. Studies showed that there were three main causes of flat threads, namely, hooked or bent pipe, out-of-round pipe coming to the threaders, and improper adjustment of the threaders themselves. This improper adjustment was isolated to one cause, i.e., improper centering of the clamping chucks.

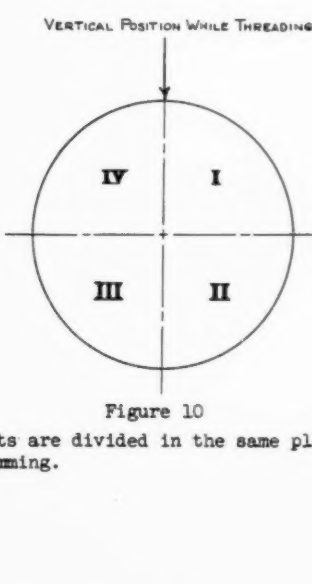


Figure 10

It was reasoned that if flat threads were generated because of the condition of the pipe coming to the threaders, the location of flat threads would occur in a random manner as far as location in relation to a fixed position of the pipe in the threaders was concerned. On the other hand, if the threaders themselves were generating the flats due to mis-alignment of the chucks, then the location of the flat threads would be concentrated non-randomly as related to a fixed position of the pipe in the threaders. Fig. 10 illustrates how the location of flat threads is determined in relation to the vertical axis through the pipe as it is being threaded. The quadrants are divided in the same planes as the chucks are adjusted by shimming.

A simplified application of the Chi Square test for randomness was developed. A work form illustrated in Fig. 11, is used so that the die man can determine whether or not flat threads are being produced in a random or non-random manner. The possibilities as shown on this work sheet are based on Chi Square at approximately 95% probability, with necessity for a decision being reached in a maximum sample size of 20 pieces.

WORK SHEET FOR STATISTICAL QUALITY CONTROL INSTALLATION NO. 514-1

Sample No. 1

Unit #6 Threader

Pipe Size & Type 4" Black

Quadrants

Date 12-15-54

Turn 8-4 Time 1:30 PM

Name F. Benk

Record Min. No. of Good Threads				
Pipe No.	Quadrant Number 1	2	3	4
1				4
2				4
3	5			
4	9			
5				7
6	4			
7				6
8	9			
9	6			
10				5
Total Pcs.*	5	0	0	5

*Count total pcs. having min. G.T. in each quadrant. DO NOT add up number of G.T.

SAMPLE SIZE = 10 Pcs.					
ADJUST Grips if any of the following sets of column totals appear (in any order)					
5	5	6 or more			
5	4	any			
0	1	any			
0	0	any			
Use bottom half and sample ten more pieces if one of the following sets of column totals appear or when in doubt.					
4	4	4	5	5	5
4	4	3	3	3	2
1	2	3	2	1	2
1	0	0	0	1	1
DO NOT ADJUST FOR ANY OTHER SET OF COLUMN TOTALS					

Note: If flat threads are about equally divided between points A & B, reduce pressure to min. and start a new sample.

Action taken 2 Shims +.0010

Figure 11

The testing plan operates by having the machine operator mark with chalk the vertical position on ten successive pieces while they are being threaded. The die-man inspects the threads on each piece and locates the quadrant in which they occur. After 10 pieces have been inspected and decision as to randomness cannot be made to 95% probability, then an additional sample of 10 pieces is marked and inspected after which a decision to adjust or not to adjust is made. It has been our experience that if the condition is not evident at 95% probability on the first ten samples, the number of flat threads is usually not enough to cause rejection at final inspection anyway.

We have found that this method of controlling threader adjustment has been one of the most effective in our whole program. Flat threads on sizes from 2 1/2" to 4", which consistently run the highest reject percentage, in November, 1954 were reduced 87% from the six months average before this plan for machine adjustment was adopted.

In conclusion, our experience has shown that on-the-job statistical controls in our pipe mills offer good opportunities for reduction of rejects, reprocessing, and consequently costs. We do not, however, make a practice of applying statistical control measures until the inherent variability of the process is known, and until we can show the Superintendent of the mill the specific benefits which he should get from such an installation. For this reason, extensive process capability studies are made before installation of statistical control techniques. We have also found that in these applications, simple, common garden variety of controls which are easily understood are more effective than more complex types of controls which no one except an S.Q.C. technician can understand.

Another plus value which has been obtained is the use of the data collected to evaluate assignable causes of variation, to isolate them and in many cases identify and eliminate them.

QUALITY CONTROL - A NEW TECHNIQUE IN THE CLOTHING INDUSTRY WITH A SHORTAGE OF SKILLED WORKERS

R.L. Murray
Hardwick Mills, Cleveland, Tennessee

"It won't work in our industry" is a common saying in industries where the skill of people would have to be charted. That craftsmen are different from machines, metals, gauges and materials is definitely true. In the clothing industry we have little trouble with machines for quality depends mostly upon the performance of human beings. Quality must be built into garments by the skill of experienced craftsmen who look upon their work as a piece of art. Therefore, for our industry we must use Quality Control in a unique way to benefit most from it.

Statistical Quality Control aids the clothing industry most during in-process operations where the human element enters in most. Of course, the incoming materials are extremely important but this can be handled easier for we are charting materials, not people. Also, knowing our out-going quality level is most important. The percent defective and the amount of sorting necessary to make shipments acceptable are cost items not to be overlooked, as well as whether we get returns or re-orders. Here again we are working with "things", not people, so we can use our proven principles of Quality Control.

All the new man-made fabrics are a challenge to the clothing industry. Dacron, Orlon, Nylon, Rayon and others require different paper patterns to "bol and must be processed differently. Here we begin to see the need for specification at the machine and some sort of control to notify management of what's going on when it happens. In other words, something to ring a bell so that supervision can take action to prevent those defective parts from going through and creating a big sorting job of the end product.

The designer or engineer and the foreman know what the specifications should be after each operation. They could write the specifications, explain them and the quality expected at each operation, then inspect and chart the operators. Without the participation of the operators in the quality effort you would create a "police-force" out of your Quality Control inspectors and get no results other than discharging some good operators.

Our experience has found it very important to get the individual operator, or operators, on the same job in "on the act" in writing the specifications. To let them explain how important their job is and how good the garments must be before and after they get them. We know that they build the quality into the product, so let them tell us what the specifications should be along with what the designer expects.

The employees have found it to be to their advantage to have the specifications written. It helps eliminate the personal "feeling" of the boss in defining "quality". Being pushed for production, he says "let it go" today and then tomorrow tries to "screw the lid down tight" on quality. As we all know, this is disturbing to all

people. It is most important to have as many measurable specifications as possible and as few attributes as possible which we cannot measure.

With the specifications written and the operator knowing he had a part in them, we need a way for the operator to show that his work is contributing to the quality product that the customer will continue to buy. This is a simple chart (C or P) at each operator's machine with control limits computed so that the outgoing quality level will be acceptable. Those operators staying within control limits should be recognized and those out of control should be worked with carefully.

When Quality Control inspectors work in this manner to help the operator and not lower his production efforts by which he is paid, the Quality Control department will have the confidence of production supervision as well as the operator. All people know that work is faster and easier to do if it comes to the operator "right".

The specification and quality expected are an aid to the Time-Study engineer in that he can tie the standard allowed minutes to the quality wanted and be fair to the operator as well as the company,

In addition to specifications for each operation, it is important to have specifications for the finished parts. In many cases, such as a finished collar, there will be measurable parts which develop from a series of attributes in the original. To treat these individual parts as completed items with a plain definition of "Go" and "No Go", minimizes a big sorting problem that can develop in the end product.

Tolerances are a very important part of any manufacturing operation. The old thinking of "exact" is not and never was possible. All people take an operating tolerance whether we admit it or not. The big problem is to find this "safe, operating tolerance" and control the operation to it. A frequency distribution chart is a sure way to get a good picture of what's going on at our critical operations and help us find reasonable and acceptable tolerances.

Using the principles described so far, we are able to divide our most difficult operations into several parts in teaching an inexperienced girl the quality needed, her tolerances and the specifications. In other words, we break down our operations for quality and simplicity as the industrial engineer breaks down and simplifies operations for production purposes. Of course, our Quality Control engineer and our Industrial engineer, working with the Designer, manage to tie our quality with our production - thus serving a double purpose. For example, one of our most difficult operations, which would ordinarily require a year's training, is sewing in a coat sleeve, due to the fact that the sleeve is $3\frac{1}{2}$" larger than the armhole. However, we produce a highly-skilled operator within a few weeks in the following manner: we tape our armhole in the old established tailoring practice, but measure with a template as we work to assure us of the proper size armhole, then we gather the $3\frac{1}{2}$" fullness in the sleeve to the pre-determined size template and the sewing in of the sleeve becomes a less skilled job. A frequency distribution chart shows us the optimum amount of fullness at the

various places of the armhole on the many types of fabrics we process. Then, the C-chart records the operator's skill.

Let's get away from generalities in Quality Control and look at some records of government contracts. On the first government contract on which we used Statistical Quality Control, we saved money in five ways, namely:

1. Reduction of inspectors
2. Reduction of supervisors
3. Eliminated repair girls
4. Reduction of material waste
5. Reduced frequent rejected lots to only one over a period of several months

We reduced our number of inspectors and finishers from 20 to 10, thus saving 400 man-hours per week. Our supervision was reduced by one-third and we put the three repair girls, as well as the other girls whose jobs were eliminated, on production.

Since that first contract we have used Quality Control successfully on several other government contracts, as well as on our commercial production of suits, sport coats, topcoats and slacks both in men's and boys' sizes.

Our application of Quality Control might be interesting to you. The Quartermaster had done a lot of work in preparing the specification and Standard Inspection Procedure on the Armed Forces clothing. We took those specifications as a base to work from on our government contracts and became so schooled in the thinking of quality at the machine that it was easy to write our in-process specifications on our commercial products. With seventy-five years' experience behind us in the manufacturing of clothing, we were able to accumulate a lot of "know-how" in written specifications.

To know and change quality levels is a very important item in a competitive product. By recording results of our old method of 100% sorting, which shows the same condition as our in-process inspection, we are able to work on our critical operations and change to a good sampling system with our 100% sorting on the defective items only. When our sampling reflects the same results as our in-process inspection, we are able to control our critical operations in the line and tighten or loosen on quality, as desired.

We know that it is better to hold a uniform product so that our customer gets what he buys, rather than make some near-perfect garments that should sell for a higher price and some that are very bad. Uniformity is one of the best things we get from Quality Control.

Inspection procedure is very important in this business of measuring the skill of people and utilizing all the skill we have available. We classify our operations into "critical", "major", and "minor" so that we will not be wasting time at the wrong place. We check our critical operations every hour and even 100% at some times at the machines. On the major jobs we take a sample twice a day and on minors we only check daily. Re-classification of operations

changes as they go out of control from in-line inspection or from results of our sampling. The in-line or in-process inspection is done in a randomized sequence, so that no one will know when his work will be inspected.

Our inspectors are rotated to minimize getting familiar or "going easy" on friends. The bundles containing defective parts are "red-ticketed" by inspectors and the supervision has them repaired at the machine or passes them by initialing the red ticket. As the sample size (n) is only 20 parts per inspection or sampling, the supervisor is responsible for having the remaining parts sorted down the line to stop defective parts from getting through.

It has been said by outstanding industrialists that a person's greatest desire is to know "how they are doing". They like to know how they "stack up" with their company. The control charts, properly maintained, will do this. The charts show our supervision where they need to work to maintain a good, quality product.

Let me summarize with this statement - Quality must be built into clothing by the skill of human hands. There must be a motivating power to keep the desire for quality alive all the time. There must be some sort of measuring device so that management can control the quality to the level they desire.

We use Statistical Quality Control, primarily, for five purposes:

1. To show us the quality of the material we buy
2. To show us the condition of these materials at each stage of process
3. To show us the Quality level in process
4. To enable us to set and change quality levels for different products
5. To assure us of uniformity of product

STATISTICAL TECHNIQUES IN RANDOM AND NON-RANDOM DISTRIBUTION OF ATTRIBUTES

Irving W. Burr
Purdue University

1. RANDOM SAMPLING AND THE RATIONAL SUBGROUP.

One of the singular contributions of Dr. Walter Shewhart was the concept of the rational subgroup. Such a sample is to be one in which all the pieces are produced and tested or inspected under as nearly identical conditions as possible. All of the variation within the subgroup is thus to be due to random causes only. Such potential assignable causes as differing personnel, material, test sets, machines or spindles, or atmospheric conditions can be let vary from one sample to another, but within any one sample they are to be held fixed. A careful record should be kept of such changes in conditions for tracing down the assignable causes. (If such a record is not kept then such detective work becomes far more difficult, if not impossible.)

If all of the pieces produced under a given set of conditions are to be included in the subgroup, then there is no problem of sampling. On the other hand if only a sample is to be tested, then there is the question of how to draw it. Now if we were truly successful in our attempt to control all conditions, then it will not matter which we choose for the sample, that is, we might as well take them from the "top of the pile." Since we really have no way of being sure that every possible assignable cause has been held fixed, however, we should be conservative and take a random sample of the product produced under the given conditions, for example those of the last half hour. A random sample is one in which each one of the pieces has the same chance to be chosen in the sample. This is the ideal to be constantly striven for. If you have in your organization someone who knows the importance of a random sample and can be counted on to select conscientiously a random sample, then you have a most valuable person. You might consider having him do all the drawing of samples for the more important job or part numbers, or tests.

All of the foregoing is applicable whether working in attributes or variables.

2. BASIC DISTRIBUTIONS FOR ATTRIBUTES, CONTROL CHARTS.

If we draw a series of random samples of n pieces each from a process which has a constant probability, p', of producing a defective piece, then the number of defective pieces, d, in the sample follows the binomial distribution, that is, the probability, P(d), of there being d defectives in the sample is given by

$$P(d) = \frac{n!}{d!(n-d)!} \quad q'^{\,n-d}\, p'^{\,d}$$

where the factorial symbol "n!" means the product of all whole numbers from 1 to n inclusive and $q' = 1 - p'$ = the probability of a good piece being produced on any one trial. The theoretical average number of defectives per sample is np' and the theoretical standard deviation of the d-values is $\sqrt{np'q'}$. Thus in a long series of samples under

statistical control, we would expect the average $\bar{d}$ and standard deviation σ_d of the number of defectives to be approximately np' and $\sqrt{np'q'}$. Thus we use control lines corresponding to np' and $np' \pm 3\sqrt{np'q'}$.

For the chart for the number of defects, c, the underlying distribution is the Poisson. According to it, if the theoretical average number of defects per sample is c', then the probability, $P(c)$, that the sample will have exactly c defects is

$$P(c) = \frac{e^{-c'}(c')^c}{c!},$$

where $e = 27183$.
The theoretical average and standard deviation of the c-values are c' and $\sqrt{c'}$. These will be approximated by a long series of c-values under controlled conditions. The control lines are thus given by c' and $c' \pm 3\sqrt{c'}$.

Now, if in a practical case, we actually do have control, then for less than 1% of the samples will the corresponding point be out of the control limits. Thus a point out under these conditions will be rare (that is, our probability of an error of the "first kind" is small). In practice there are two things which come into the problem. In the first place, we usually do not know the true values, p' or c', and must therefore use in their place the observed averages $\bar{p}$ or $\bar{c}$. This gives us only approximate limits. The second disturbance is that often there are assignable causes present. Thus p' or c' is not constant from sample to sample, but instead varies. Does this mean that the binomial or Poisson distribution is no longer applicable and that our control limits are wrong? In a sense it does but what we are doing is testing the hypothesis that we have just one population, either binomial or Poisson. A point outside the control limits throws doubt on this hypothesis and indicates (subject to some risk) that while that sample was produced and tested there was a different population at work, say a binomial with a higher or lower p'. We want to find out the cause for that shift and so seek among the possible assignable causes to find what was different at that time than at others.

If no point is outside such control limits then we say the hypothesis is tenable or permissible, that is, all of the observed variation among the points can perfectly well be due to chance alone. The process is "in control."

3. DISTURBANCE, CHANGES IN POPULATION.

As we have just seen, one thing which can happen is that the population is not fixed but jumps around, that is, p' or c' is varying from sample to sample. What effect does this have upon the chart? It gives greater variability, that is, σ_d' or σ_c' is greater. This result is due to Lexis. In plain language if p' varies around some $\bar{p}'$, then the points vary more and tend to go outside the limits based upon $\bar{p}'$. The greater p' varies, the greater this tendency to show lack of control. The control chart is perfectly all right here, because it is aimed at catching just such changes in p' or c'. The only real danger is that in analyzing past data a few high c- or p-values actually produced under an assignable cause, may so increase $\bar{c}$ or $\bar{p}$ that these points are below the upper control limit. This unfortunate tendency is fairly well held in check by our conventional method of setting the width of the control

band from $\bar{c}$ or $\bar{p}$ rather than σ_c or σ_p. The latter would be much more influenced by extremes than is $\bar{c}$ or $\bar{p}$.

Another possibility is for the sample to be stratified. Thus, for example, we can take one piece from each of twenty heads or molds. It could be that all twenty behave alike, but more often than not there will be at least some tendency for some heads or molds to average more defectives or defects than the others. In this case a sample of one from each will be a stratified sample. (In one extreme case the four corner cavities of 16 almost invariably produced defectives and the others did relatively well. To the suggestion that the corner cavities not be loaded, the foreman replied that they could not stand the 25% loss in production!)

The net effect of such stratification in a sample, is to <u>decrease</u> the variation among the points. If the variation among p' for the twenty heads or orifices is slight then the cut in σ'_p will be little and there will simply be less chance for a point to go out than for unstratified samples. It will take a slightly stronger assignable cause in say, material, to show out. On the other hand, if there is larger variation among the twenty, then the points may be much more clustered around the center line, than for the random sample. This may well show up on the chart by the limits seeming to be far too wide to give any chance of going out. Thus stratification may lead to "too good" control. Charts with the points running too close to the center line might well be suspected of coming from stratified sampling.

A third disturbance is that in which we pool a large amount of data for a management chart. Thus 100,000 or 1,000,000 piston rings or caps may be a single day's production. If, for example, $\bar{p}$ is running at .025, then what are the control limits for p for a day's production?

$$.025 \pm 3\sqrt{\frac{.025 \times .975}{1,000,000}} = .025 \pm .00047$$

$$= .02453,\ .02547.$$

Now anyone familiar with such overall quality figures knows that it is fantastic to hope for the daily production figure to lie within such extremely narrow limits. Is the mathematics wrong? No, it is simply that we have samples from a great number of populations of varying p''s and the collection varies from day to day. In such a host of raw material, part numbers, processing and inspection we are bound to have many assignable causes. Such is inevitable and what we should seek in this one, daily figure, is <u>not</u> assignable causes, but instead a "super assignable cause", that <u>is</u>, one which affects the whole shop. To make such a chart we can analyze our series of p-values just as though they were measurements X. They can be handled by $\bar{X}$ and R charts or by an individuals chart for X's and a moving range. This same technigue can be used in cases of heavy stratification within samples or in case of spoilage of bulk product, such as, surface rejection of steel plate or proportion of wire defective.

A fourth way in which the basic distribution (binomial or Poisson) may be upset is through lack of independence from piece to piece. Grant [1] gives an example in which 2300 rubber belts were made in a mold at one time. The fraction defective for a sample from, or 100% inspection of, product from a single mold will not follow the binomial.

It will exhibit more variation, because if one belt is bad, the chance of another being bad is greater. Bad rubber could cause most of the belts to be bad, or good rubber most to be good. Lack of independence causes the binomial to be inapplicable. Here again we can treat a series of p-values just as though they were measurements X.

Another viewpoint, which is not distinct from the foregoing is that of the so-called "contagious" distributions. The general idea is that, for example, in the case of defects on a sheet of paper, the occurrence of one defect tends to enhance the chance of another defect. This is the contagion; the defects are not independent. In this sense then we have a correlating tendency as in the preceding paragraph. If the average number of defects is relatively small, then this will tend to show up in a frequency tabulation by there being too many occurrences of zero defects and too few occurrences of one defect for an ordinary Poisson distribution having the same mean $\bar{c}$. Accident - proneness is one common example of this type of influence. Many have no accidents, and although many have one accident, there are fewer than for a pure Poisson, and there are a few with a larger number of accidents than would be expected in a Poisson.

Now such a tendency in a frequency distribution does not necessary imply such "contagion", because this can also be the result of non-homogeneous distributions. Thus if we have samples from a mixture of two or more Poisson populations we tend to have a frequency tabulation exhibiting the same tendency as just described. Thus if sheets are drawn from a mixture of sheets half of them from $c' = 1$ and half of them $c' = 5$, the average is 3 defects, but there will be more cases of $c = 0$ and less of $c = 1$ than for a Poisson distribution with $c' = 3$. An excellent mathematical discussion of this subject of contagious distributions is given by Feller [2].

4. CHECKING RANDOMNESS.

Control charts give a natural way to test the randomness of a series of observations, for example, of defective vs. good in order: G G G D G G D D G, etc., or of the numbers of defects per unit: 0 2 1 1 2 1 2 0 0 2, etc. We simply subgroup such a string of observations into samples of any desired number of pieces. In the first case we might take the first 10 or 20 pieces in a single sample which yields a fraction defective. Then we form a p chart, or in the case of defects, we can use a $\bar{c}$ chart. If there are strings of defectives or defects, such charts will show this up.

In checking the randomness of sampling from a lot we can well take our main sample of perhaps 200 pieces in 10 samples of 20 pieces each, the first 20 being the first subsample, etc. A control chart of these 10 samples should show control if the sampling was at random, even if the lot was not produced under controlled conditions. Of course this is not true for defects charting if the control has been extremely bad, so that, for example, most units have about 2 defects each, while a few have 50 or so. But there should still be control no matter how wildly out-of-control the process was, if using a fraction defective chart. The reason is that, to start with, the lot has just so many defectives, whether these were produced all at once or piecemeal. Random sampling with equal chance for each piece will not tend to get too few nor too many in any case.

Another good way in which to check for randomness of order is to count the number of runs of defective pieces and of good ones. (This method is useful for fraction defective problems and not for defects counts.) For example, we might have the following order G G G G G D D G G G G G G D G G G G D D D G G G G G G G G D D G G G G D D D D G G G G G D D D G G G D G G G G D D D G G G G G G D. Here we have 45 good pieces and 20 defectives. The number of runs in this set of 65 pieces is 18. Is this what we should expect from chance? It looks less than average, since the defectives have tended to bunch. But is this tendency significant? We have the following for the mean and standard deviation of the number of runs, U [3]:

$$\overline{U}' = \frac{2\,d\,g}{d + g} + 1$$

$$\sigma_U' = \sqrt{\frac{2\,dg\,(2\,dg - d - g)}{(d + g)^2\,(d + g - 1)}},$$

where g is the number of good pieces, and d of defectives. We may assume normality if both d and g are at least 20. In our example $d = 20$, $g = 45$ so we have

$$\overline{U}' = \frac{2 * 20 * 45}{20 + 45} + 1 = 27.69$$

$$\sigma_U' = \sqrt{\frac{2 \cdot 20 \cdot 45\,(2 \cdot 20 \cdot 45 - 20 - 45}{(20 + 45)^2\,(20 + 45 - 1)}}$$

$$= 3.398$$

Now suppose that we use the 5% significance level (1 tail test). Then we want to know whether the probability of 18 or fewer runs is .05 or less. We want to find the probability of 18 or less and include the "18 block" as in a histogram of bars. This block runs from 17.5 to 18.5. The standard score is therefore

$$\frac{18.5 - 27.69}{3.398} = -\,2.70.$$

A normal curve table gives the probability as .0035. So the observed 18 runs is significantly below the expected, and hence there is a significant tendency for the defectives to bunch.

If there are fewer than 20 defectives in a set of data we can resort to the quite extensive tables given by Swed and Eisenhart [4]. Selections of these tables are reproduced in a text by Dixon and Massey [5]. There are other tests of randomness of order, such as those on the distribution of lengths of runs and on the length of the longest run. One is cautioned, however, not to apply too many tests of randomness, since every series of numbers, no matter how obtained, will have some peculiarity which could make at least one of the infinitude of possible tests show "significant non-randomness."

5. USE OF CHI-SQUARE AND CONTINGENCY TABLES.

The technique called chi-square is discussed in most textbooks. It is a technique for analyzing observed frequencies against given hypotheses, usually some form of independence. Westman and Freeman [6] give an example in which the relation between gas holes and no gas holes in castings is compared with three classes of percent carbon: under 1.155, 1.155 to 1.195 and over 1.195. In the data analyzed there was no evidence of a relationship. Some good examples of the use of chi-square and a frequency distribution showing a contagious tendency are given in an article by Gore [7]. It is possible to check a series of samples for attributes which are three or more categories. For example, if in tin-plating sheets could be called good or defective, then we could run a fraction defective chart. But if they are called "good", "menders" or "waste wasters," then we have 3 classes or categories in each sample of say 112 or 1120 sheets. We can check control as a whole by means of chi-square.

There are many tricky points to be watched in studies by chi-square. A good presentation of many of these are by Lewis and Burke [8]. There are several good articles on chi-square in a recent number of <u>Biometrics</u> [9].

6. REFERENCES

1. E. L. Grant, <u>Statistical Quality Control</u>, McGraw-Hill, New York, 1952, p. 260.
2. W. Feller, "On a General Class of 'Contagious' Distributions," <u>Annals of Mathematical Statistics</u>, Vol. 14, No. 4, (Dec., 1943), p. 389.
3. A. Wald and J. Wolfowitz, "On a Test Whether Two Samples are from the same Population," <u>Annals of Mathematical Statistics</u>, Vol. 11, No. 2, (June, 1940), p. 151.
4. Frieda S. Swed and C. Eisenhart, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," <u>Annals of Mathematical Statistics</u>, Vol. 14, No. 1, (Mar., 1943), p. 66.
5. W. J. Dixon and F. J. Massey, Jr., <u>Introduction to Statistical Analysis</u>, McGraw-Hill, New York, 1951.
6. A. E. R. Westman and R. W. S. Freeman,"Statistical Control of the Manufacture of Managanese Steel Tank Shoes," <u>Canadian Metals and Metallurgical Industries</u>, Vol. 8, (June, 1945), p. 38.
7. W. L. Gore, "Statistical Methods in Plastics Research and Development," <u>Industrial Quality Control</u>, Vol. 4, No. 2, (Sept., 1947), p. 5.
8. Don Lewis and C. J. Burke, "The Use and Misuse of the Chi-square Test," <u>Psychological Bulletin</u>, Vol. 46, No. 6, (Nov., 1949), p. 433.
9. <u>Biometrics</u>, Vol. 10, No. 4, (Dec., 1954).

THE USE OF RANGE CHARTS

Eugene C. Yehle
School of Business Administration
University of Michigan

"The control of uniformity" in products, processes, and measurements characterizes, as well as any single phrase, the objective of statistical quality control in the chemical industry. This paper directs itself toward the general problem of uniformity and to the discussion of appropriate applications of range and difference charts as control devices.

To be controlled, uniformity must first be defined and measured. There is no doubt, for example, that the A particles in Table 1 are more uniform than the B particles, but the question is how much more uniform?

TABLE 1
Weights of Individual Granules of a Molding Powder

Grams	Type A	Type B
.0120 - .0139		卌 ////
.0140 - .0159		卌 卌 ////
.0160 - .0179	/	卌 卌 //
.0180 - .0199	卌 卌 //	卌 //
.0200 - .0219	卌 卌 卌 卌 /	///
.0220 - .0239	卌 卌 /	///
.0240 - .0259	卌	//

The apparent extreme weights for A are .0160 to .0259 grams and for B are .0120 to .0259 grams, or ranges of .0099 and .0139 grams, respectively. The ratio of these ranges (.0099/.0139), gives an index of relative variability, and it might be said that A is 71 per cent as variable as B.

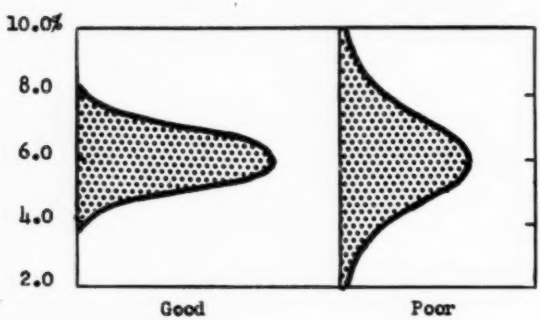
Note, however, that if the single A particle in the class .0160 to .0179 had been absent, the ratio would have been .0079/.0139 or 57 per cent. This substantial change (from 71 to 57 per cent) illustrates the instability of ranges in large samples and suggests the need for a better measure of uniformity, or its opposite--dispersion.

The best and most frequently used measure of dispersion is the standard deviation--defined as $\sigma = \sqrt{\Sigma(x-\bar{x})^2/N}$ where $\bar{X} = \Sigma X/N$, the arithmetic mean. For the data of Table 1, σ_A = .00187 g. and σ_B = .00320 g. giving a ratio of 58 per cent, compared to 71 per cent obtained from the ranges. Omitting the smallest A particle reduces σ_A to .00184--a trifling change which illustrates the greater stability of the standard deviation over the range for cases involving moderately large amounts of data.

Where measurements become available at periodic intervals, as in routine control checks, the aggregate number of observations may be large, but the number available at any one time is usually small. In these cases, the superiority of the standard deviation is not particularlarly great, and the fundamental purpose of analysis changes from making a single comparison between two distributions, to making routine, repetitive comparisons to a standard.

An example concerning the within batch variation of impurities in a compound will illustrate the point. Historical records indicated some periods when the impurities averaged 6 per cent with a standard deviation of .7 per cent. In other periods, although the average remained about the same, the standard deviation was noticeably larger--about 1.5 per cent. The within batch distributions for these two periods are shown in idealized form in Chart 1.

CHART 1
Distributions of Per Cent Impurities within Batches from Two Production Runs



Batch-by-batch control required an answer to the deceptively simple question: "Is this batch uniform?" If a sufficiently large number of samples are taken from different parts of the batch, the answer becomes obvious: the calculated σ is either near .7 or near 1.5 per cent. However, if only a limited number of samples are drawn, and this is the most usual case, the answer will not be obvious. In fact, one must be satisfied with a rather indefinite answer, such as: "It is more probable that this set of samples represents a batch typical of $\sigma' = .7$ than of $\sigma' = 1.5$ per cent, or vice versa." (σ' stands for the theoretical or population standard deviation, and should be distinguished from σ which stands for the standard deviation of a set of sample measurements. σ' and σ may be quite different numerically for small sets of data, but they become more nearly equal as the number of measurements increases.)

THEORETICAL CONSIDERATIONS:

Since the decision with respect to uniformity requires an inference based on the uncertain evidence of a small number of analyses, the behavior of samples must be understood. Three observations may be made in this connection:

1. The sample standard deviation (σ) tends to be smaller than the population standard deviation (σ').

2. The sample range (R) is almost as reliable as σ for estimating σ' from a small group of samples.

3. Both R and σ fluctuate considerably from one group of samples to the next even though each group comes from the same population. Maximum probable limits for such fluctuations are known, however, and these limits form the basis for control decisions.

The solid line in Chart 2, portraying the average relationship between σ and σ' for various subgroup sizes, explains more fully the first observation noted above. The standard deviation computed from two observations ($n = 2$) tends to be only 56 per cent as great as the actual population standard deviation, but this ratio increases quite rapidly and for $n = 10$ it becomes 92 per cent. The ratios in Chart 2, of course, are the standard c_2 factors used in statistical quality control, and are derived from the sampling distribution of the standard deviation (8) assuming a normal population. Failure to recognize the fact that sample σ's tend to be somewhat small will naturally result in overoptimistic estimates of uniformity.

The fact that the sample range can be used to estimate σ' is due to work by Tippett (10) and others. (See (6) for derivation.) As in the case of estimating σ' from σ, the average ratio of R to σ' must be known. The dashed line in Chart 2 demonstrates that this ratio also depends upon the number of observations. Thus, while R tends to be 1.1 times as large as σ' when $n = 2$, for $n = 10$ R becomes 3.1 times as great on the average for normal populations. These plotted ratios are the well-known d_2 factors.

But in order for R to be as useful as σ, it must not only give good estimates in the long run; it must be as dependable as σ in the short run. (See observation two above.) The efficiencies in Table 2 (1) show that, while R is not quite as reliable a short-run estimator as σ, for sample sizes up to $n = 10$ it is almost as good. In practical application, this means that one can take advantage of the ease of calculating R and use it in preference to σ for many problems of uniformity control.

CHART 2
Dispersion Ratios for
Various Subgroup Sizes

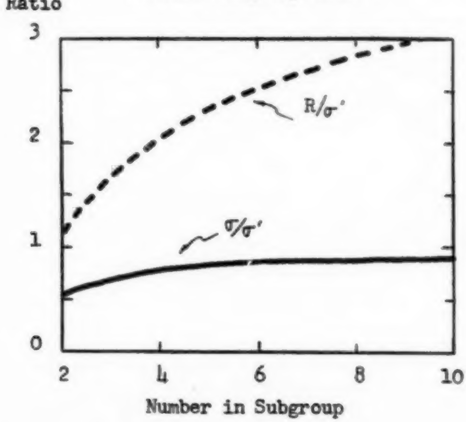
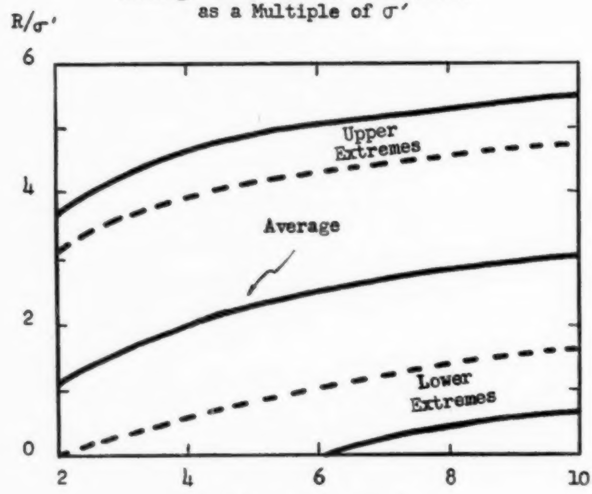


TABLE 2
Efficiency of R Relative to σ
for Estimating σ'

Subgroup Size	Efficiency
2	1.00
4	.98
6	.93
8	.89
10	.85

One final point remains: the limits within which R might vary due to chance alone--that is, in repeated random samplings under a constant set of conditions. The outer limits in Chart 3 represent three sigma limits based on the usual D_3 and D_4 factors. For n = 4, R averages 2.1 times σ', but in any single instance R/σ' may be as low as zero or as high as 4.7. The inner dashed lines of Chart 3, representing 95 per cent probability limits (4), permit a somewhat more definite statement to be given. Thus for n = 4, the odds are 19 to 1 that the sample range will vary between .6 to 4.0 times the population standard deviation.

CHART 3
Average Value and Extremes for R
as a Multiple of σ'

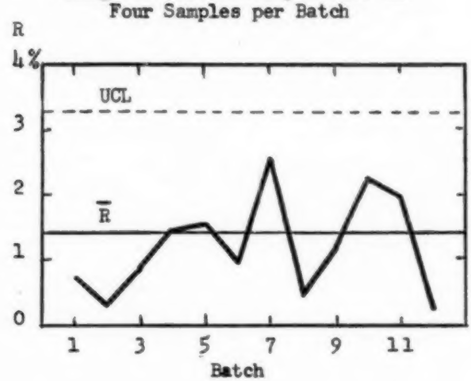


The considerable volatility of R (and σ) in small samples is disturbing. It shows that highly reliable estimates of σ' cannot be made unless more data are available. On the other hand, the actual problem in control does not really require a good estimate of uniformity,

but rather requires a decision such as the following: the evidence is insufficient (or sufficient) to conclude that the uniformity has changed significantly since the preceding sample was taken.

Consider the preceding illustration on percentage impurities in a compound. During acceptably uniform operating periods σ' is .7 per cent. In practice, this means that if samples are regularly taken from each of the four centrifuge loads that constitute a batch, the range of per cent impurities in the four samples will average 1.44 points ($d_2\sigma'=2.06 \times .7$). Furthermore, the three sigma upper control limit will be 3.3 points ($D_4 \bar{R} = 2.28 \times 1.44$ or $D_2\sigma' = 4.7 \times .7$ from Chart 5), and the lower limit will be zero. As long as the ranges from the four wheelcake samples average 1.44 per cent over a period of time and fluctuate between 0 and 3.3 points, it is appropriate to conclude that the within batch variation is acceptably low. Chart 4 shows the behavior of the ranges from a series of batches.

CHART 4
Range of Per Cent Impurities in
Four Samples per Batch



The actual ability of the range chart to detect shifts in σ' depends upon several factors including the magnitude of the shift, the number of samples in a group, and the type of control limit in use--three sigma limits, 95 per cent probability limits, etc. In the case under discussion, if σ' changes from .7 to 1.5 per cent, an out-of-control range will occur about 4 times in 10. This probability clearly is not very high--a fact which raises numerous questions that are discussed at some length by Scheffé (9). Narrower control limits or a larger number of sample observations or probability center lines may help to remedy the situation, but in the final analysis, it must be recognized that sample data can never guarantee perfect information. Nevertheless, some data, though meager, plotted at regular intervals does provide a useful index for judging the presence or absence of control in an industrial situation.

APPLICATIONS:

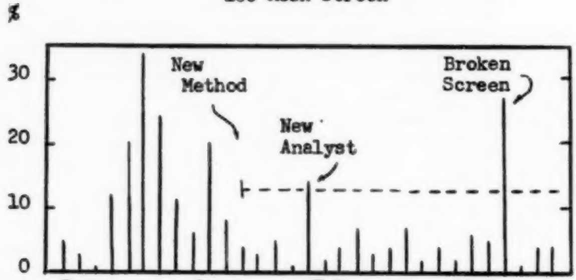
So much for theory. As frequently noted, chemical applications might well begin in the laboratory. Only when the reliability or reproducibility of analyses has been established can the quality control analyst "go hunting" for process and product variations.

The range chart is particularly suitable for measuring laboratory precision. It can be used, as illustrated in Chart 5, for comparing initial test results with rechecks on the same sample. If the sample identity for rechecks is concealed, a bona fide estimate of laboratory precision will result. Thus, as long as the chart shows a state of control with an average range of 4.0 points, the standard deviation of analyses is about 3.5 points ($\bar{R}/d_2 = 4.0/1.13$). This means that if 30 per cent of the sample supply of material is finer than 100 mesh, the laboratory will report something between 23 and 37 per cent 19 times in 20 ($\pm 2\sigma'$ limits). Consequently, values of 25 per cent for one batch and 35 for another do not invalidate an assumption that both batches possibly have the same percentage of "fines". This is particularly true since the magnitude of another potential source of variation--the sampling process--has not been stated.

Chart 5 could also have been plotted as a difference chart (12). The zero line would be the initial analysis and the recheck would be plotted as a plus or minus range, depending upon whether it was higher or lower than the first result. The control limits of the standard range chart may still be used, although the actual sigma level of these limits is not quite the same for the two types of charts. Indicating the direction of range gives added flexibility in analysis by helping to point out trends and biases that may be traceable to changes in atmospheric conditions, deterioration of reagents and batteries, and fouling of equipment. It should be noted, in connection with difference and range charts, that while laboratories frequently report the average of several measurements rather than just a single observation, this practice will not invalidate nor complicate the use of charts as long as the averages are routinely based on the same number of measurements. (See (5) and (1) for other applications.)

Where the range is obtained from more than two results, the difference chart does not apply. However, some of its benefits can still be obtained by identifying the highest and lowest observations directly on the chart. Had this been done in Chart 4 it would have been apparent that the sample from the fourth wheel tended to be high and the one from

CHART 5
Range between Checks on Per Cent Through
100-Mesh Screen

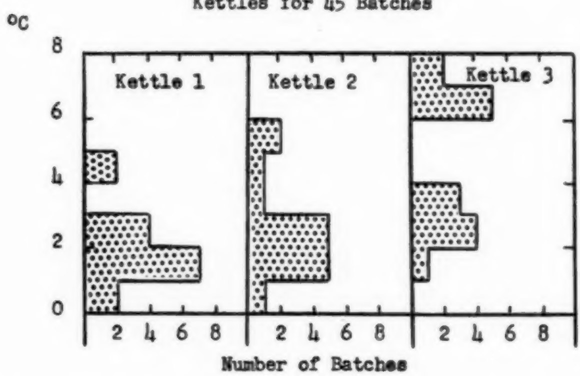


the first wheel tended to be low. Thus there was a kind of built-in variability that the control chart itself did not detect.

The high degree of automatic instrument controls common to the chemical industry at times seems to make statistical controls superfluous. However, the intricate instrument networks themselves require watching, and the obvious thing to do is to control the controls statistically. Here again the range chart finds many useful jobs.

Where uniform temperature control is important, the range of high to low temperature during an 8 hour shift or during the processing of a single batch should be plotted. Such a chart will quickly reveal worn linkage in control mechanisms, inadequate supplies of cooling water or steam, corrosion deposits, and even carelessness on the part of operators who assume that the controls do all of their work for them. See Chart 6.

CHART 6
Temperature Range in Three Kettles for 45 Batches



Pressure ranges can be handled in similar fashion, as can weigh tank scale errors, and pumping and processing times. Where there are several temperature and pressure points the maximum variation within the entire system can be plotted. (See Bicking (7) for other suggestions.)

Control charts on product quality characteristics and product yields present the most difficult problems of all. Viscosity, refractive index, per cent impurities, per cent chlorine, or any of innumerable measurements often fluctuate appreciably. True, the fluctuation may not be "real" in the sense that it is due to lack of perfect precision in laboratory measurement or to a heterogeneous product from which it is difficult to obtain a representative sample. (In the latter case a satisfactory solution first requires a definition of product quality.) But supposing that the characteristic measured really possesses a serious batch-to-batch or time-to-time component of variation, what can be done about it? The quality control department must do more than announce the fact; it must help to trace down the cause. The range charts on the product can do this only if they are tied in with charts on the process and on the raw materials. "Tied in" fundamentally means choosing rational subgroups--a subject well discussed by Grant (4).

In the case illustrated by Chart 6, batch viscosities were also related to kettle number, and high viscosity variability appeared coincident with high temperature variability. Many relationships are far more complex, however, due to the appreciable time lag between the operation of a causal factor and the final measurement of its effect in the product. To make matters worse, there is usually a sequence of possible causal factors whose composite effect is reflected in the product only hours or days later. Statistical methods of analysis can not possibly sort them out unless considerable ingenuity has been used to synchronize control charts all along the line. In addition, more frequent sampling at strategic intervals <u>during</u> processing is usually inevitable.

Tracking down sources of variation ultimately leads to some experimentation. Here range charts are useful too, although they serve primarily to evaluate rather than to control the experiment. Suppose, for example, that data are available for several batches from each one of a group of reactors, and the object is to look for significant differences among the reactors. Variance analysis methods handle this problem best, but rough answers may be obtained by employing the range. (See (3) for many illustrations.)

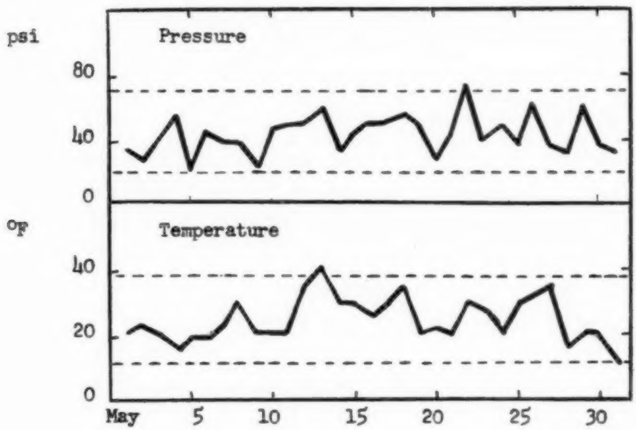
A first step is to obtain a measure of the experimental error. The average range for the replications within a reactor will serve this purpose, and at the same time the behavior of the range chart will indicate the validity of an important assumption for further analysis--that the within-reactor error variances are homogeneous. For the latter purpose, 5 per cent probability limits, common in research work, are more suitable than the D_4R or three-sigma levels used for process control.

The next step is to compare the reactor means. Tukey's method (11) of arranging the means according to size and testing the differences between adjacent ones provides essentially a moving range chart that is useful for a preliminary study. The chart limits are based on the $\bar{R}$ of the first step--the within-reactor variation. For an approximate 5 per cent significance level the limits will be $2\sqrt{2}(\bar{R}/d_2\sqrt{n})$, where n is the number of replications for each reactor. Out-of-control points divide the results into groups for further analysis.

This use of the range, while crude, facilitates rapid analyses and simple graphic presentations of results. Other research applications appear when several sources of variation must be evaluated. Steam flow, for example, depends upon both temperature and pressure. These quantities fluctuate considerably, and the question arises as to their relative importance in providing reasonably precise estimates of net steam consumption from readings on metered steam. Chart 7 shows that the daily ranges are substantially in control and, therefore, that it is reasonable to assume a constant variability of temperature and pressure even though their average levels can and do fluctuate significantly. Given the estimates of σ' based on $\bar{R}/d_2$, and the multipliers for temperature and pressure from steam tables, standard formulas for combining variances (2) can be applied. They show that the two variables are almost equally important in affecting the reliability of the required correction factors, and that no extra care or equipment are necessarily required for pressure as contrasted with temperature readings.

CHART 7

24-Hour Range of Pressure and Temperature
For a Steam Turbine



In conclusion, range charts represent an exceedingly flexible tool of analysis in the chemical industry where close control is a by-word. If Research invents a workable process, and Engineering develops adequate facilities, the quality control function is largely one of helping Production maintain uniform operating conditions. This may require control of laboratory precision and special evaluation studies as well, but even in these areas the range, as a quantitative expression of uniformity, has many applications.

REFERENCES

(1) C. A. Bennett and N. L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, Inc., 1954, p. 165.

(2) O. L. Davies, Statistical Methods in Research and Production, Oliver and Boyd, London, 1949, p. 44.

(3) A. J. Duncan, Quality Control and Industrial Statistics, Richard D. Irwin, Inc., Homewood, Illinois, 1953.

(4) E. L. Grant, Statistical Quality Control, McGraw-Hill Book Company, New York, 2nd Edition, 1952, pp. 159, 186.

(5) F. E. Grubbs, "The Difference Control Chart with an Example of Its Use," Industrial Quality Control, July, 1946, p. 22.

(6) A. Hald, Statistical Theory with Engineering Applications, John Wiley and Sons, Inc., New York, 1952, p. 319.

(7) J. M. Juran, Quality Control Handbook, McGraw-Hill Book Company, New York, 1951. (Section 11, C. A. Bicking, "Quality Control in the Chemical Process Industries," p. 561.)

(8) J. F. Kenney, Mathematics of Statistics, D. Van Nostrand Company, New York, 1939, Part 2, p. 135.

(9) H. Scheffé, "Operating Characteristics of Average and Range Charts," Industrial Quality Control, May, 1949, p. 13.

(10) L. H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika, 17, 1925, p. 364.

(11) J. W. Tukey, "Comparing Individual Means in the Analysis of Variance," Biometrics, June, 1949, p. 99.

(12) E. C. Yehle, "Laboratory Precision and Specification Limits," Analytical Chemistry, July, 1953, p. 1047.

QUALITY CONTROL IN THE PRODUCTION OF ALUMINUM FOIL

O. H. Bishop
Reynolds Metals Company

All industry desires control of quality in products and recognizes its importance for success, yet there is probably more variation in administrative thinking with regard to its planning than all other essentials combined. Within any assembly of a group of professed "quality guardians", we might expect to find all extremes. One may consider 100% final inspection most reliable, even though performed by any employee not possessing the required skill of an operator. Another might revolutionize an entire organization with such an elaborate system that production is burdened with its execution. In either event, costs are prohibitive, the program folds up, production returns to reliance on good breaks of luck and Management gropes for new suggestions. Somewhere between these extremes is the ideal plan, one, which with patience and care, may be tailored to apply to an individual company; there is no general plan. Perhaps then a description of our Company's program may be of interest. It will be related by following the manufacture of one product in each step of process and illustrating the methods used by concentrating on one characteristic. For simplicity we will select gauge control, because it is variable and measurable in numbers.

In the search for the "ideal plan" we seized every convenient opportunity to take advantage of the vast amount of information that the ASQC and its associates make available for Industry. We must admit, however, it seemed difficult to apply the principles of statistics to continuous coils of aluminum, because they do not conveniently lend themselves to, pardon the use of the overworked phrase, "nuts and bolts". There is usually, though, at least one product which might encompass the general pattern of operations and yet be convenient to sampling. The one we chose for this purpose is Household Foil.

1. It is a controlled run-of-the-mill product which in final form (25 foot lengths) may be conveniently sampled by any suggested plan.

2. Defects, ordinarily not significant in the starting aluminum may be very pronounced in thin foil gauge.

3. In production, it crosses the boundaries of four interplant activities.

4. Its cost of production must be reasonable to the budget of our most bargain seeking consumer, the American housewife.

5. In like manner, she is most demanding in quality for her money. Note -- Mrs. Housewife does not keep a micrometer within her collection of kitchen gadgets, but she can notice gauge variations by the number of times she can smooth out a piece and reuse it.

6. This critic exercises her right to voice her opinion, does so by personal letter and is often outspoken in doing so.

In lecture form a series of 3D pictures or photographs of the operation brings a close-up of the "highway" (1) of travel of our product. The

reprints offer a glimpse of the highspots, one of each interplant activity -- mining and reduction of the aluminum -- breakdown of an ingot in heavy gauge rolling -- finish rolling and spooling -- finally, convertor spooling and packaging as a finished product. Past experience has proven that quality suffers its most severe blows during manufacture in crossing the interplant barriers. In our over-all Company organization these are under separate Divisional Management. In one-plant operations they might be departmental barriers. In either event 100% inspection is impossible in the semi-finished product; we can only see the outer turns of the coil of aluminum. If some method could be devised for semi-finished-goods inspection by a supplying plant and acceptance inspection by a receiving plant, it would be economically prohibitive and would require elaborate systems besides aggravating problems of disposition of rejected material between plants or departments.

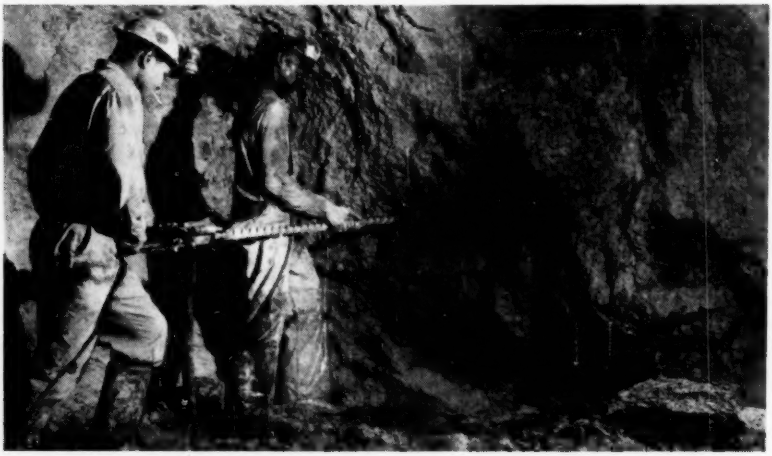


Fig. 1. Underground Mining of Bauxite.

Most disturbing to all production men is any action on the part of Inspection which results in a tie-up of much needed material for an indefinite period, or the withholding of cores (or mandrels) when a shortage probably already exists. His production record slumps.

Disagreements on the severity of defects are common, especially when interplant transfer is made between individual plants. The primary reason is a lack of sympathy for the other fellow's trouble. Misunderstandings in the nomenclature of types of rejects runs a close second.

Eventually, inventory control sounds the gong and so the disturbance has grown to a proportion which can only be arbitrated on managerial level and discord prevails.

We might consider this situation as Dr. A. V. Feigenbaum so ably expressed it (2):

"Modern Quality Control integrates the usually unco-ordinated approaches to control of quality into an over-all program for

a factory. Quality Control activities, like 'Topsy', have just growed' during the past decades. The value of an over-all, co-ordinated plan in place of sprawling, disjointed activities is well known in factory administration."

We, too, had long recognized that Quality Control is basically the co-ordination of inter-departmental activities concerned with the manufacture of a product and this will continue to be our AIM, but we seemed to lack the proper tools or effective systems for measuring and pinpointing trouble. In reviewing our process we observe some time-proven but "disjointed activities" in the same sequence of our "photographs".

1. Chemical composition of the aluminum must conform to requirements for uniform rolling characteristics both hot and cold.

2. All mills are equipped with departure gauges which are continuous in operation. The deflection of a galvanometer needle informs the operator when the sheet travels "off gauge."

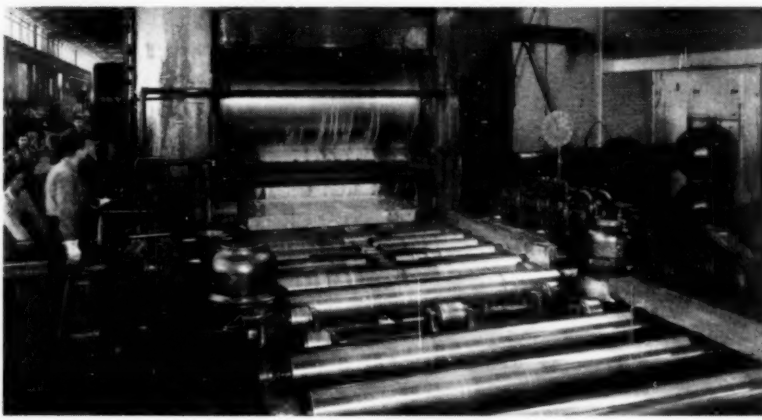


Fig. 2. Hot Mill -- Ingot Breakdown.

3. Additional physical inspection in hot rolling is a control measured by micrometer.

4. In finish rolling in the heavy gauges down to .0017" a micrometer is satisfactory on the coil ends inspection.

5. During final passes, plant checks on trims or cut-outs aid in immediate control, but final inspection from finishing spooling is made by obtaining the weight of the finished roll and the yards contained measured by production meter. By use of a chart designed to consider width of web, the number of square inches per pound are reported and termed "yield".

6. In convertor spooling, sampling is varied by necessity for

vigilance, but with consideration for consumption of items of finished product. Here, actual spot testing is employed, that is, a 3" x 3" area is weighed on a laboratory analytical balance and by table is converted to both yield and corresponding gauge.

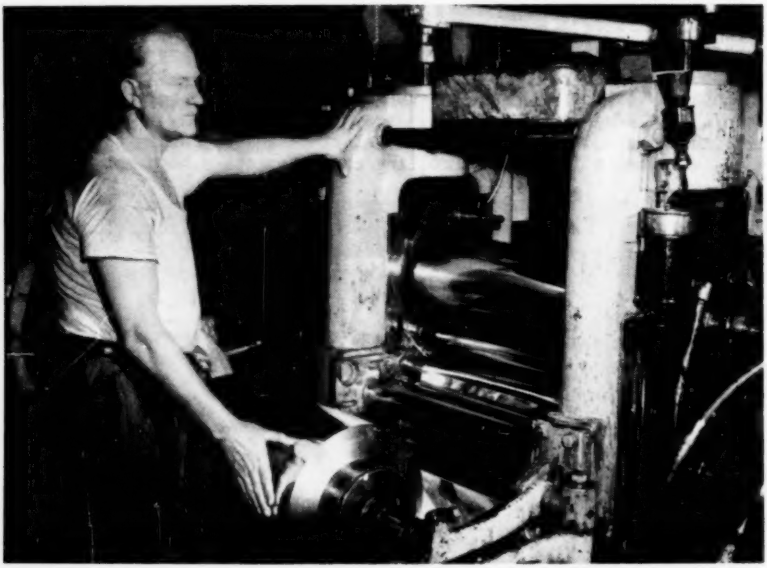


Fig. 3. Foil Finishing Mill.

These methods succeeded in building up immense quantities of remote cabinet records with selected ones used only when useful to bolster an argument that a doubtful, semi-finished product is worthy of subsequent processing. Let's recall now a reprint in our collection of literature, "Fundamentals of Quality Control" by Dr. Lloyd A. Knowler. (1) "Much material has been collected and filed away in a desk drawer where it does no one any good." — "biased data should not be secured to support a whim. The collection, the tabulation, the analysis and the interpretation of the data tie together, and when properly united are profitable to the company and all concerned." Perhaps then, here is a tool or system that we have long sought, which will lead toward a realization of our AIM.

Dr. Knowler suggests, "In the development of a quality control program two important factors should be considered. First, get started, but do not go too fast. The second important factor is to understand the fundamental principles." In presenting a parallel to the control chart it is "likened to a highway." "The control chart is thus a picture or a photograph." "Personnel must see the success of the plan and be carried along with it."

In our "picture" or "photograph" presentation of the "highway" of travel of our product, attention was invited to the crossing of troublesome interplant barriers. Our decision was to tackle these intervals,

considering their co-ordination as being most vital to the success of the venture, consequently a good place to "GET STARTED". The interpretation of the advice, "but do not go too fast," obviously means do not attempt a revolution. A tactful approach is to recognize that many good, sound measures of control have been well established since the industry's beginning and there are many persons in influential positions now who are very proud of their initial accomplishments.

Fig. 4. Convertor Spooling and Packaging.

Which ones then of the previous listed "disjointed activities" may be useful and wherein lies our authority to revise our system of compiling the data to be effective toward correction? Again we resort to our store of advice. When questioned regarding his opinion of the one most important factor necessary for the success of an effective control program, Dr. Paul J. Mundie (3) replied,"Secure a sponsor in top echelon and solicit his confidence and support." In this instance we are most fortunate that our General Manufacturing Manager compiled our previous principles of "time proven" thought into an amazingly inclusive letter of authority and assigned a central office position of Quality Control Manager. Our written program is composed to preserve continuity of our AIM, since the development of all principles are backed by reference to our original letter of authority. Time does not permit a detailed description of our individual plant organization for the quality control function except to state that they are of uniform pattern under direction of a Plant Quality Control Manager who is actually authorized as an extension assistant of the general staff function, namely, Operations Quality Control. This affords streamlined contact between plant and staff functions with a ready and able assistance offered to all other general departments, within the plant, interplant and staff offices.

Thus our authority is clear cut,now for a means of compiling without too great an increase in personnel. Our production control panel has been in use many years. It represents as efficient, modern method of collect-

ing data from all operating machines. Each is wired with an operator-to-panel call system. Its records afford data suitable to compile severity and frequency of defects which have interrupted production. Roving inspectors assist the operators in decisions of rejects. Simultaneously, pertinent records of production such as pounds of aluminum processed,down time, etc., are recorded. Tentative weekly estimates of trends are teletyped to the plant supplying the semi-processed aluminum input. At the end of each month, form R-853-2 is compiled by the plant Quality Control Manager using summations of the classified data and supported by actual inventory of pounds of rejects set aside for return, each portion labeled as to reason. This action constitutes a practice to remove defects as they occur to reduce cost of needless, subsequent processing of defective material and, of course,prevent them going through into the finished product.

In this limited time we are unable to describe details required to iron out all of the kinks. However, rejects are reduced, because we strip out defective portions rather than return complete coils. It may actually be declared as a system of sorting,bearing in mind that incoming coils must be unwound to determine what is inside. This sorting, however, has furnished classified data, which our Staff Statistician has constantly used to determine significant facts for corrective action and to present charts as trends for Management observation. To date, 46 such charts are currently compiled. Form R-853-2 is becoming to be known as a "Quality Analysis Report". Disposition of the rejected portions is simultaneously arranged by designing a letter of transmittal to the report with reference to a long established General Manufacturing Circular 01:01 (Interplant Credit and Cost Exchange). Only slight revisions to it and a complete distribution list insures automatic clearances.

Periodic meetings of individual Plant Quality Control personnel encourages sympathy with the other fellow's problem. A binder of actual defect illustrations created a uniform nomenclature in discussions for corrective action. When dealing with interplant transfer we now encounter no more difficulty in following the same "highway" than the modern motorist contends with in following U. S. Route #1 from the bottom of Florida to the tip of Maine. Some central government commission, however, first must have linked up a maize of separate state roadways and lent aid to straighten out the crooks and helped design a route which could serve the most for the least mileage expended. Standard road maps resulted to insure the correct choice of the proper route. In like manner, an authorized central control body, equipped with a written program furnishes a constant guide of authority in the development of our control methods and their applications.

Now let's reconnoiter, we have succeeded in winning a good degree of confidence, and created a showing of considerable savings in dollars by reducing costs of useless subsequent processing. To this point Top Management patience has been quieted temporarily, because there is evidence that something is certainly "clicking". Let's not relax though and live on these laurels. The real showing in savings of dollars remains in, not merely arresting defective material at intervals of interplant transfer, but in correcting the cause at the source. Quality Analysis Reports are beginning to "lay bare many previously remote cabinet records." Our system as described up to this phase is dependent upon a "whip" in the form of penalty of reject as determined by a form of acceptance transfer. Our Managers' patience will not remain quiet long;front line supervision will then resort to cover up, passing the buck or preferably, of course, tak-

ing corrective action. If our Department, then, is equipped to stimulate more corrective action by being "more than just quality watch dogs", when "Quality Control gets a checkup" (4), then we need not fear our future.

In the parallel to our "highway", we have constantly consulted the authority of our roadmap and selected a proper route. In the construction of our interplant highway we have not fought the natural contour of the land; if a gap was too deep, we did not fill, we bridged it. When impractical to cut or tunnel, we took the shortest course around a long standing hill. In such events, we used the best of the maize of state roads whenever possible, because originally they were mapped out to serve the best interest of that state and some Senator in Congress looks after its welfare. Dr. Knowler, however, expressed a more complete meaning in his relation to a highway. He described the "pavement" to correspond to the "safety zone"; along either side is a "shoulder" or "caution zone" and, of course, the "ditch", the "danger zone". "The boundary between the shoulders and ditch represents the specification limits." Intent on this meaning and desiring to stimulate corrective action by assisting front line supervision in placing numerical values on manufacturing capabilities, we turn to our Staff Statisticians compilation of Quality Analysis Reports. Control charts will furnish each one concerned with a "picture or a photograph" of his part of the "highway".

The natural approach in our product, Household Foil, is to chart the gauge (or yield) by sampling from the convertor spoolers. It was very quickly determined that four or five measurements made from one 25-foot length did not yield the correct data to plot an average and range chart. This is probably best explained by reasoning that the range of variation in gauge does not occur within 25 feet of a roll of foil containing some 5,000 yards. Our sampling plan is best suited by securing one measurement from each of 4 or 5 of the 25-foot lengths produced from one feed roll to the spooler. These tests are made by weighing on a laboratory balance, 3" x 3" squares, cut by template. They are, of course, "spot checks" or "individuals". Figure 5 represents one such chart made several years ago.

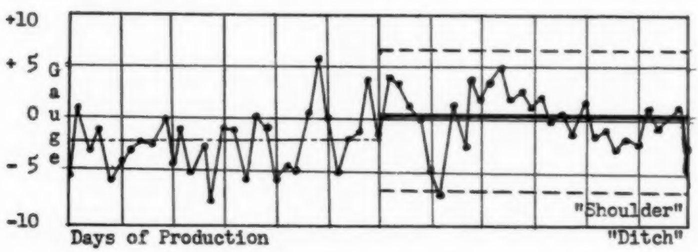


Fig. 5. $\bar{X}$ (Average Chart.

In sale, no guaranteed gauge range is declared; therefore arbitrary numbers, multiples of 5 are used to represent variations.

In setting the stage for this initial demonstration we were careful to choose locations where the finishing pass mills and convertor spoolers were under the same Manager's jurisdiction. It was indeed encouraging when, of his own accord, this Manager inquired, "Where were these samples taken?" When we replied, "On the convertor spooler," he said, "Let's get them over on the finishing mill where the operator can do something about

ing data from all operating machines. Each is wired with an operator-to-panel call system. Its records afford data suitable to compile severity and frequency of defects which have interrupted production. Roving inspectors assist the operators in decisions of rejects. Simultaneously, pertinent records of production such as pounds of aluminum processed,down time, etc., are recorded. Tentative weekly estimates of trends are teletyped to the plant supplying the semi-processed aluminum input. At the end of each month, form R-853-2 is compiled by the plant Quality Control Manager using summations of the classified data and supported by actual inventory of pounds of rejects set aside for return, each portion labeled as to reason. This action constitutes a practice to remove defects as they occur to reduce cost of needless, subsequent processing of defective material and, of course,prevent them going through into the finished product.

In this limited time we are unable to describe details required to iron out all of the kinks. However, rejects are reduced, because we strip out defective portions rather than return complete coils. It may actually be declared as a system of sorting,bearing in mind that incoming coils must be unwound to determine what is inside. This sorting, however, has furnished classified data, which our Staff Statistician has constantly used to determine significant facts for corrective action and to present charts as trends for Management observation. To date, 46 such charts are currently compiled. Form R-853-2 is becoming to be known as a "Quality Analysis Report". Disposition of the rejected portions is simultaneously arranged by designing a letter of transmittal to the report with reference to a long established General Manufacturing Circular 01:01 (Interplant Credit and Cost Exchange). Only slight revisions to it and a complete distribution list insures automatic clearances.

Periodic meetings of individual Plant Quality Control personnel encourages sympathy with the other fellow's problem. A binder of actual defect illustrations created a uniform nomenclature in discussions for corrective action. When dealing with interplant transfer we now encounter no more difficulty in following the same "highway" than the modern motorist contends with in following U. S. Route #1 from the bottom of Florida to the tip of Maine. Some central government commission, however, first must have linked up a maize of separate state roadways and lent aid to straighten out the crooks and helped design a route which could serve the most for the least mileage expended. Standard road maps resulted to insure the correct choice of the proper route. In like manner, an authorized central control body, equipped with a written program furnishes a constant guide of authority in the development of our control methods and their applications.

Now let's reconnoiter, we have succeeded in winning a good degree of confidence, and created a showing of considerable savings in dollars by reducing costs of useless subsequent processing. To this point Top Management patience has been quieted temporarily, because there is evidence that something is certainly "clicking". Let's not relax though and live on these laurels. The real showing in savings of dollars remains in, not merely arresting defective material at intervals of interplant transfer, but in correcting the cause at the source. Quality Analysis Reports are beginning to "lay bare many previously remote cabinet records." Our system as described up to this phase is dependent upon a "whip" in the form of penalty of reject as determined by a form of acceptance transfer. Our Managers' patience will not remain quiet long;front line supervision will then resort to cover up, passing the buck or preferably, of course, tak-

ing corrective action. If our Department, then, is equipped to stimulate more corrective action by being "more than just quality watch dogs", when "Quality Control gets a checkup" (4), then we need not fear our future.

In the parallel to our "highway", we have constantly consulted the authority of our roadmap and selected a proper route. In the construction of our interplant highway we have not fought the natural contour of the land; if a gap was too deep, we did not fill, we bridged it. When impractical to cut or tunnel, we took the shortest course around a long standing hill. In such events, we used the best of the maize of state roads whenever possible, because originally they were mapped out to serve the best interest of that state and some Senator in Congress looks after its welfare. Dr. Knowler, however, expressed a more complete meaning in his relation to a highway. He described the "pavement" to correspond to the "safety zone"; along either side is a "shoulder" or "caution zone" and, of course, the "ditch", the "danger zone". "The boundary between the shoulders and ditch represents the specification limits." Intent on this meaning and desiring to stimulate corrective action by assisting front line supervision in placing numerical values on manufacturing capabilities, we turn to our Staff Statisticians compilation of Quality Analysis Reports. Control charts will furnish each one concerned with a "picture or a photograph" of his part of the "highway".

The natural approach in our product, Household Foil, is to chart the gauge (or yield) by sampling from the convertor spoolers. It was very quickly determined that four or five measurements made from one 25-foot length did not yield the correct data to plot an average and range chart. This is probably best explained by reasoning that the range of variation in gauge does not occur within 25 feet of a roll of foil containing some 5,000 yards. Our sampling plan is best suited by securing one measurement from each of 4 or 5 of the 25-foot lengths produced from one feed roll to the spooler. These tests are made by weighing on a laboratory balance, 3" x 3" squares, cut by template. They are, of course, "spot checks" or "individuals". Figure 5 represents one such chart made several years ago.

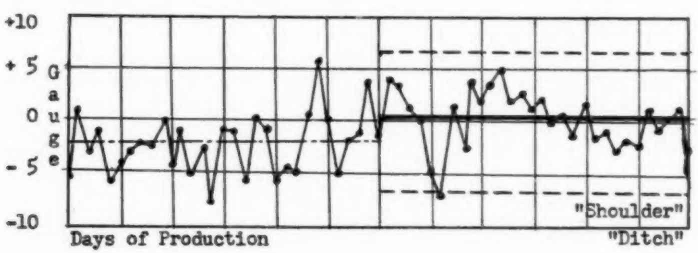


Fig. 5. $\bar{X}$ (Average Chart.

In sale, no guaranteed gauge range is declared; therefore arbitrary numbers, multiples of 5 are used to represent variations.

In setting the stage for this initial demonstration we were careful to choose locations where the finishing pass mills and convertor spoolers were under the same Manager's jurisdiction. It was indeed encouraging when, of his own accord, this Manager inquired, "Where were these samples taken?" When we replied, "On the convertor spooler," he said, "Let's get them over on the finishing mill where the operator can do something about

it," and added, "even if it requires stopping a mill to obtain a sample." Now,you realize an entire paper might be written on the subject of "stopping a mill" alone so we trust you will consider the fact that once given such authority, discretion must be exercised in devising a practical procedure. This paper, as you note, does not propose to describe details, but to express principles in developing a program.

Step one was immediately accomplished while back-tracking down the "highway", setting our boundaries, or if we may, "laying a uniform width of modern paving," and "photographed" by constructing a control chart.

We mentioned that during finish rolling in the heavy gauges down to .0017" a micrometer is satisfactory for coil ends inspection. Obviously, however, if an inspector is "on his toes" and does locate "off gauge" it is a simple matter to strip those coils until the reading shows "within specs.", before releasing. Not even intimating that any of our operators would even consider such, but one could turn out an excellent record in pounds of production if he would roll to the heavy side, then finish near or "on the button". Many deceptive coils would pass the coil end inspection. Our statistics, however, you recall were based on "spot testing" the coil throughout its length; therefore, will not tolerate deceit. To make a long story short, this corrective action has pressed its way down the "highway" to its origin. Our continuous operating departure gauges are now equipped with recording time charts,which in themselves construct beautiful range charts.

Finally to convince our own Sales Department of the accuracy of our present plant controls, we solicited their aid to request various salesmen all over the country to purchase individual rolls of Household Foil in the same manner as a customer would. This is done periodically in lots of 30 to 50 rolls. Figure 6 is self-explanatory. Our Vu-graph presentation allows us to superimpose any chart from plant control over the composite Market Sample Histogram to study similarity. We realize that technically there are many factors to consider when comparing the relations, but we also know that the value of this demonstration has been very effective in winning confidence in our methods.

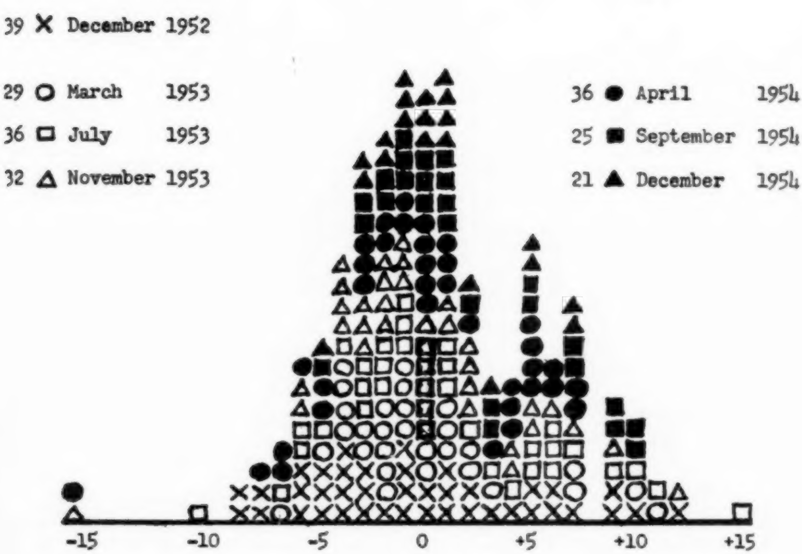
In summary, we are convinced that we have followed the full meaning of Dr. Knowler's advice. As illustrated, we feel confident that our system is so designed as a result of "not going too fast" that we could handle the situation now even if someone did order "three or four carloads". Other defects such as pinholes, surface blemishes, etc. are handled in a similar manner. The pattern established in this one product induces the desire for the same improvement in all of our operations.

Dr. Burr suggests, (5) "Use your imagination! Keep on studying."

Our concise files of statistics begin to afford sound backing to the granting of authority to our Quality Control Department. Individual Plant Managers are becoming firmly convinced that they have streamlined contact with ones who are sympathetic to their troubles in adjacent plants. Line supervisors and operators have become more confident each day of the manner in which guidance is spotlighted and are more aware that they are the ones who manufacture quality.

Geographical Distribution of Market Sampling

Recommendations made to the Sales Department in regard to the variation in population are only estimated with consideration for convenience; therefore, do not propose to represent precise sampling.



Statistical Distribution of Results of Gauge Determinations

References

1. "Fundamentals of Quality Control", Lloyd A. Knowler, State University of Iowa. Reprint, Industrial Quality Control, July, 1946.

2. "Quality Control. Principles, Practice and Administration", A. V. Feigenbaum.

3. "The Human Factor in Quality Control", Paul J. Mundie. Sixth Annual ASQC Convention, Syracuse, N. Y.

4. "Quality Control Gets a Checkup", Report in Business Week, December 18, 1954. Conference of 250 Quality Control Men at Illinois Institute of Technology.

5. Notes, "Advanced Course in Statistical Methods", Irving W. Burr, Purdue University.

QUALITY CONTROL TECHNIQUES IN AIRCRAFT ELECTRICAL WIRING SYSTEMS

Frank H. Howard
Fairchild Aircraft Division
Fairchild Engine & Airplane Corporation

This presentation will deal with the airframe manufacturers' quality problems in the electrical wiring he installs to accomodate electrical, radio and electronic equipment. In the main, the Quality Control problems are common, and the control techniques are applicable to all aircraft manufacturing plants.

For those of you who may not be familiar with an airframe manufacturer's electrical assembly operations and tests, let's take a short tour of a typical plant.

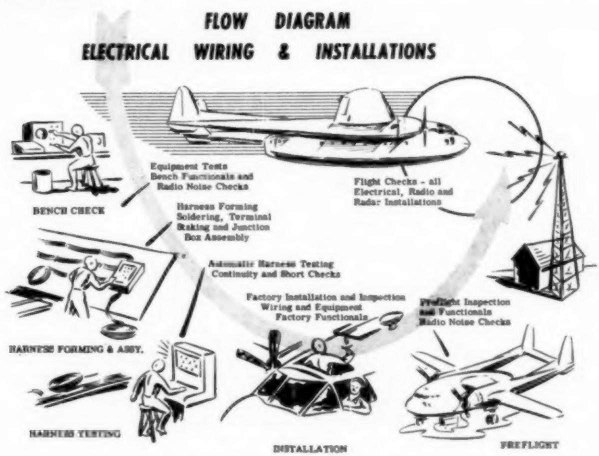


Figure (1) - Flow Diagram

First, the wire harnesses are made by laying pre-cut wire on forming boards, tying them together and soldering on connector plugs or staking or crimping solderless terminals to the wire ends. After inspection, these harnesses are either routed directly for assembly or are installed in electrical junction boxes on the bench. The harnesses and boxes are then installed in the various airplanes sub-assemblies, which, in turn, are mated into a complete airplane.

At the same time in another part of the plant, the radio, radar, electronic and electrical equipment are being pre-installation tested for performance to specifications. Radio noise and other tests are made at this point. The units are then routed to join the airplane at the various pre-planned assembly stages.

A complete functional test is performed on all systems in final assembly. This includes operation of the inter-phone system, operation of transmitter control heads to check for proper frequency selection, use of test equipment on radar sets, testing of the electrical landing gear controls, operation of the lights, operation of the engine cowl flaps, etc.

The factory completed airplane receives additional electrical tests during ground engine run, where radio contacts are established with the ground station, radio noise tests are made on the whole airplane, radio compass compensation is checked, and other similar type tests are accomplished.

The very process of flight involves the use of many electrical and radio systems but some special flight patterns are required for radio-altimeter checks, radio-compass tests, marker beacon operation, etc.

ELECTRICAL BENCH OPERATIONS - CONTROLS

Constant surveillance must be given to basic electrical manufacturing operations, i.e., wire stripping, soldering, crimping and staking. Malpractices in these daily repeated operations can cause serious problems in aircraft electrical systems. Here are safeguard techniques used to control potential failures which originate during bench operations:

Soldering

To control aircraft soldering, there is substantiation for the belief that a program should be conducted before workers are permitted to solder. The program would provide:

1. An eye test.
2. An intensified education in the process.
3. A practical training (worker would solder sample joints representing conditions of actual manufacture).
4. A qualification-type test.

While assurance that the solderer has the required ability is a "must", more important, perhaps, is the necessity for constant surveillance of his daily work. The most competent solderer can produce poor work when he is physically fatigued or when his mind is occupied with trying personal problems.

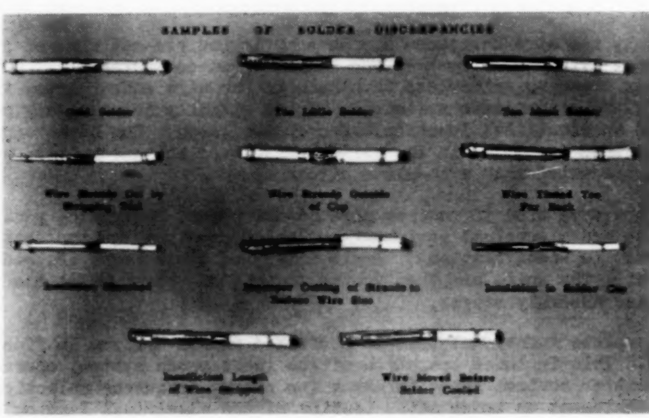


Figure (2) - Samples of Soldering Discrepancies

To control the soldering quality, aircraft plants use:

1. Visual inspection.
2. Periodic sampling of production joints (laboratory sectioning and examination).
3. High amperage, millivolt tests (sampling).
4. "Pull-Tests" of joints (sampling).
5. High potential tests (most commonly used on Co-Axial Cables).

It should be noted that Government specifications do not make mandatory all of these solder tests, and that the extent of testing, over and above the Government specifications, varies considerably in the airframe plants.

Quality Control people in our industry view with interest methods now being advanced to reduce the "operator's skill" variables. By unique design application, manufacturers of electrical connector plugs are making a progressive effort in this field. Cups into which wires are inserted for soldering contain the proper amount of solder. The cup is heated by the worker who inserts a tinned wire and holds the wire until the solder cools and the joint is fixed.

Under laboratory control, tests have been accomplished to prove the effectiveness of this design. Deliberate malpractices were used in the preparation of the test samples as follows:

1. Wire not tinned prior to cup insertion.
2. Wire not inserted far enough into cup.
3. Wire pushed into cup bottom with excessive pressure.
4. Wire wiggled during cooling cycle.
5. Solder poured out of cup.
6. Wire ends cut on bias.
7. Wire strands not connected (30% cut away).

SAMPLE	CURRENT			PULL TEST	TYPE OF DELIBERATE MALPRACTICE USED FOR EXPERIMENTAL PURPOSE
BENDIX	135 AMPS	200 AMPS	270 AMPS	BREAK PT. LBS.	
	VOLTAGE DROP MV				
I	7.5	11.0	15.0	207.0	No tin on wire
II	8.0	13.0	20.0	*	No tin - pulled up
III	8.0	11.0	16.0	478.0	No tin - excessive rosin
IV	6.0	9.0	12.0	1042.0	Tinned - pushed down excessively
V	6.5	9.5	15.0	1034.0	Tinned - wiggled until cool
VI	6.5	9.0	14.0	922.0	Tinned - pulled up very high
VII	7.5	11.0	15.0	154.0	Tinned - solder poured out of cup
VIII	7.5	12.0	16.0	268.0	Untinned - 30% strands not connected
PERFECT SAMPLE	5.5	8.0	12.0	1104.0 Wire Parted	
WIRE ALONE	6.0	9.0	12.5	1104.0	
AMPHENOL #I	5.5	8.0	12.0	*	
AMPHENOL #II	5.5	8.5	12.0	1164.0 Wire Parted	
	* CROSS-SECTIONED				

Figure (3) - Test Data

The laboratory sectioned sample joints for examination in the polished and etched condition, conducted high amperage millivolt tests on soldered wire to cup samples and made tension tests on other soldered wire to cup samples. The laboratory results indicated that, in spite of the deliberate effort to produce poor results, the millivolt drop was negligible and the pull test values which were wide-spread, were, in several cases, much higher than anticipated. Efforts on the part of designers to reduce the penalty for factory human-element sub-quality workmanship should be applauded.

Another recent design which, if adopted, may reduce "operator skill" variables, eliminates soldering by using a taper pin crimped to the wire. The pin is driven into the connector (cannon plug receptacle) by use of a tool. The joint is electrically and mechanically sound and replacement of wires, using the same tool, is a simple operation. In this process, the single soldered joint is replaced by two (2) mechanical joints (wire crimped to tapered pin and pin driven into connector receptacles). Obviously the purchase of tapered pins, special receptacles and assembly tools are required if this design is adopted.

Terminal Staking, Crimping

Power and automatic operated equipment is used in all aircraft plants for the staking and crimping of wires or cables to terminals. This equipment reduces the "human element" error. Hand operated tools must be used when assembly rework is required. Experienced personnel are important to staking or crimping operations. Extreme care must be exercised in the case of tools for aluminum terminals.

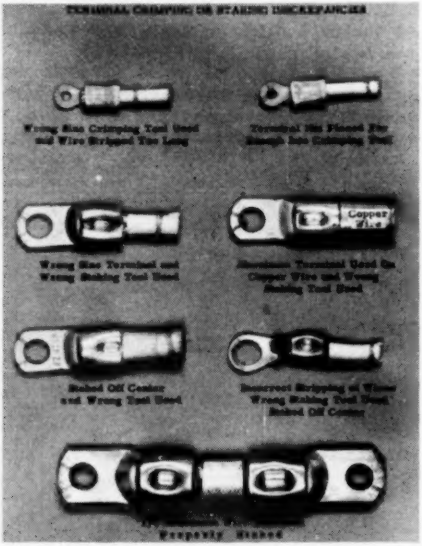


Figure (4) - Examples of Terminal Crimping and Staking

To control the quality of staked or crimped joints, the following tests are made:

1. Periodic inspection of tools or machines.
2. Visual inspection (including wire stripping).
3. Periodic tests (sample basis) of joints produced by each worker consisting of:

   a. Pull tests.
   b. Dimensional measurement of stake indentation.
   c. Presence of petroleum oxide-inhibitor (aluminum terminals).
   d. Section examination (laboratory).

Automatic Wire Harness Continuity Testing

The use of automatic ring-out machines falls into two general categories. One where the harness to be tested has connector plugs on both ends. This ring-out can be accomplished very rapidly by plugging into the machine and making one automatic run. In the other category, the harness has a connector plug at one end, and terminals in a junction box or control panel at the other end. This requires a more complicated tester. In this case, a circuit may have as many as 6 to 8 ends, and pass through switches or relays. This ring-out test involves connecting the panel or box and wiring to the test machine and proceeding through the test, operating the necessary switches and relays as the test progresses. "Go-No-Go" principles are used and light signals flash for acceptance or rejection.

Figure (5) - Automatic Wire Harness Continuity Tester

Automatic harness testers have made a tremendous contribution in the reduction of inspection man-hours, over the old pin-to-pin plug ringout method. A marked contribution is also apparent in the quality of electrical systems when the automatic harness testers are used. The incidence of trouble which would require rework at installation has been substantially reduced by this technique.

Potting of Connector Plugs

Some aircraft plants are potting electrical connectors to assure freedom from corrosion caused by water or moisture at soldered joints. Other companies have studied or are studying the potting process. The compound used in the process is synthetic rubber, which sets up to the consistency of a pencil eraser.

Figure (6) - Potted Connector

Here are some of the factors that are being considered by the Industry:

1. Corrosion resistance is provided.
2. Moisture proofing is provided.
3. Dirt, chips, filings, etc., are eliminated.
4. The process provides a 27% average saving in weight per plug.
5. Irvolite sleeves can be eliminated.
6. The process requires special equipment (pressure extruders, upright storage racks, cure ovens, etc.)
7. The system is inflexible to quick engineering changes, to the handling in cases of shortage of connectors and to normal rework.
8. The process is somewhat messy and time consuming.

ELECTRICAL INSTALLATIONS - CONTROLS

As electrical wiring installations are made in the airplane, potential malfunctions may result. Sound design principles and good shop practice; or, if you will, "Inspection Control characteristics" are necessary to prevent:

1. Wire chafing (routing problems, poor slack distribution).
2. Cannon plug corrosion (absence of drain holes, drip loops, etc.)
3. Grouping of critical circuit wiring with other cable runs.
4. Incorrect wire gage, insulation and clamps.
5. Incorrect wire lengths.

In addition to the routine floor inspection control, it has been found that in combat of these offenders, several control techniques are being used to great advantage in our industry.

Mock-Up

The "mock-up" system prior to the wiring of the first production airplane is in use in many plants. Through adequate trial and error installations in the "mock-up", many potential electrical wiring problems can be eliminated. The mock-up provides a practical means for determining wire routing and lengths. Quality Control can assure that the final quality concepts are incorporated at the time the design is still fluid.

The mock-up affords, in addition, an opportunity for those who look ahead to equipment changes. Realistic design provisions at the mock-up stage can eliminate problems and assure high quality installations for future production.

Some companies mock-up an airplane's electrical power system and distribution circuits, and move the entire mock-up into an extreme temperature altitude chamber for realistic simulation of flight testing.

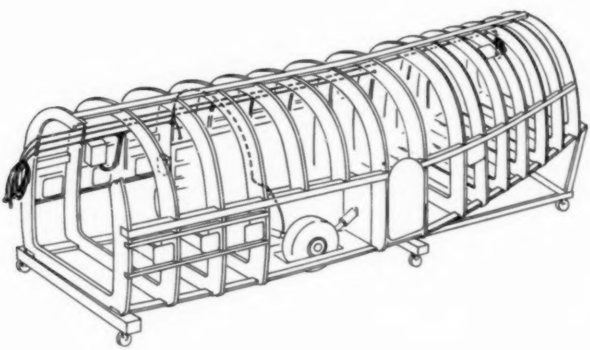


Figure (7) - Wiring Mock-Up

Wiring Team

Another technique which is being used to advantage is the "wiring team" idea. The "wiring team" is usually comprised of electrical personnel of the Quality Control, Engineering and Service Departments. An important function of the team is to work out details of wire routing and wire lengths; which is not practical to do on the drafting board. In many ways, the teamwork of such a group can be invaluable to the goal of manufacturing aircraft with first quality electrical systems. The Service Department representative contributes by using his knowledge of the Using Agency's base maintenance problems. The Quality Control representative can assure that quality considerations are included and can also accomplish invaluable liaison with Inspection and Production personnel. The Engineering representative, in coordinating the drawing data, can assign to the drawing detailers, call-out responsibilities for clamp locations, clamp sizes, points requiring insulating sleeves, etc.

The wiring team works on a "first installation" either on a mock-up, an experimental airplane, or a "pilot production model". The team then periodically reviews production aircraft. Problems occasioned by changes, due to equipment type replacement, can be eliminated by action of the team. Surely from a Quality Control viewpoint, a team ever critical and determined to refine can pay material dividends.

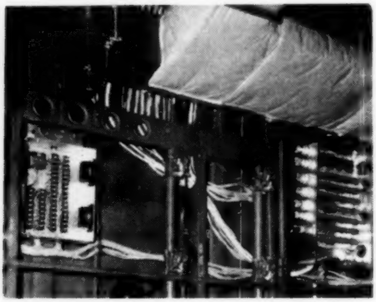
Wire Length Miscalculation Costly

A recognized problem, which involves waste wiring, interferes with the normal flow of work and can jeopardize quality, is the failure of bench assembly people to cut wiring to proper length prior to its installation in the airplane. In such cases, additional lengths are allowed and are cut to the proper dimension by the installation people. Often, in these cases, the hand staking of terminals to the wire ends is not equal in quality to the machine staking job normally accomplished on the bench. The determination of wire lengths is an Engineering responsibility; the control of the wire length problem is in the hands of the installation inspector. While it is natural for us to assume that "it can't happen in our facility", with the frequent wiring changes taking place, it may pay for us to have a look at this potential problem and, if it exists, we should take steps to assure the prior planning required for its elimination.

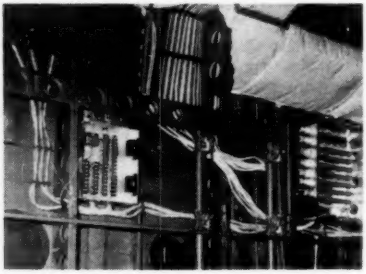
Photographs and Photo-Touch-Up

Many plants use photographs to depict details of installational wiring, tieing, routing, protecting, clamping, etc., to simplify requirements and assure uniformity between installations. Junction box detail arrangement photographs are quite popular in this application.

A technique employed by at least one (1) airframe manufacturer involves the use of re-touched photographs to depict wire routing, clamp locations, etc., in complicated areas of the airplane. Through a photo-touch-up process, aircraft structure is removed in these areas to expose the complete electrical wiring details. The photos are used by installing workmen and by Inspectors. This plan has reduced the production manhours utilized in installation and has improved the quality of the electrical systems through method standardization which curbs human-element errors.



Before Touch-Up



After Touch-Up

Figure (8) - Photo-Touch-Up

Aluminum Wire in Aircraft

The use of aluminum wire is important because of weight saving and because aluminum is more readily available than copper. Aircraft plants have experienced quality control problems in dealing with aluminum wire. Here are some of the problems:

1. Inadvertent use of copper staking tools on aluminum terminals.
2. Precision measurement requirement for staked dimensions.
3. Loss of oxide-inhibitor during staking.

Engineering considerations in the use of aluminum cables involve restriction of the terminals from vibrating equipment and from excessive heat (either external source or from adjacent connections that can transmit heat to the terminals).

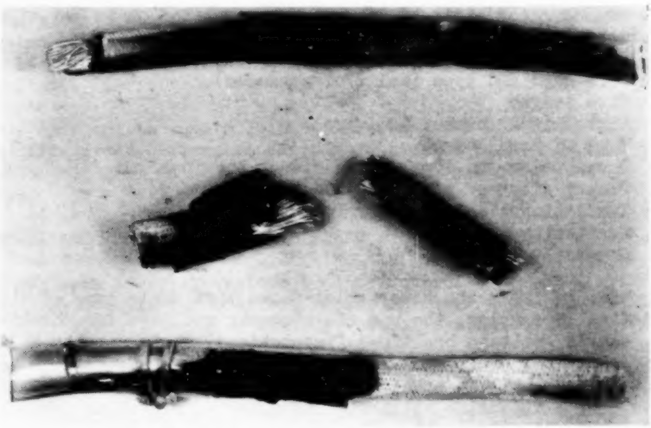


Figure (9) - Burned Aluminum Cable Connectors

Burning at aluminum connectors is usually due to a high resistance joint. Resistance increase is caused by: (1) the relaxing of a staked joint (cold flow effect in aluminum) to the point that the intimate contact between the wire and connector is lost, (2) the loss of intimate contact because of the difference in the thermal coefficient of expansion between aluminum and other metals, (3) the build-up of non-conducting oxide which takes place on aluminum surfaces when exposed to air as explained in (1) and (2).

Statistical Quality Control Applications

It can readily be seen that aircraft electrical wiring processes are applicable to Statistical Controls in many forms because:

1. There are many characteristics to be inspected.
2. There are many people conducting the operations, which would make the use of $\bar{X}$ and R charts practical.
3. There are many separate processes that can be controlled, i.e., stripping, soldering and crimping.

4. There is enough similarity between many of the parts produced to permit grouping of lots and characteristics.
5. The quality requirements are similar and therefore only one or two AQL's need be used, simplifying sampling requirements.

Normally, on bench operations such as soldering, crimping and part numbering, a form of statistical process control is established, whereas on the actual inspection of assembled wire harnesses, a product type of inspection is desired. To take a typical case of the latter, to illustrate possible statistical controls, consider a lot of wire harnesses made up of many wires of different lengths, gages, insulation, etc., assembled to various connectors by soldering, crimping or staking. Tool control and periodic sampling of wire stripping tools, crimping and staking tools and soldering equipment plays an important part in this process control. Knowing that the processes are under statistical control, we set up check lists to inspect the completed units for other characteristics, such as:

1. Wire gage and type (solid or stranded)
2. Wire length
3. Wire circuit number (on wire)
4. Type of insulation
5. Proper size and type of individual terminal
6. Proper size and type of connector plugs
7. Insulating sleeves on terminals
8. Continuity
9. Shorts

A review of these characteristics shows that while some of them may be statistically sampled, the use of automatic machines for checking continuity, shorts, voltage breakdown (insulation), etc., permits a 100% inspection fast enough to meet production schedules and, because machines are used, results in 100% assurance. The other characteristics are sampled to a single AQL using attribute inspection. Variables inspection is normally not applicable to these inspections because of the lack of dimensional criteria and the necessary measuring equipment. The use of acceptance visual standards, as noted herein, forms the basis for acceptance criteria.

CONCLUSION

The importance of electrical systems in today's aircraft makes the Quality Control of these systems a vitally serious and essential control. As service experience has accumulated, both the Government and the airframe contractors have changed their requirements to provide the ultimate in dependability and serviceability of electrical systems. If we are to be on our toes as Quality Control people, we must employ the most efficient of the known control techniques and we must conduct a search for more efficient techniques on a day-to-day basis.

QUALITY CONTROL BUDGET METHODS

Paul E. Allen
Beech Aircraft Corporation

GENERAL

The budgeting of all departments is one stage or function of many good managements in achieving their general mission of administering the activities of the company in an efficient manner so that equitable profits will result. Actually the profit a company makes year after year is the major basis by which the owners and stockholders can judge the efficiency of a company's administration.

The various industrial managements have many tools such as: operating ratios, departmental reports, historical records, comparative business indexes, organization reports, etc., which are used as a guide in aiding them in directing their companies to the realization of an equitable profit.

In my opinion, an equitable budget system is a stimulant that encourages effective departmental administration and is a challenge to departmental managers in the development of more efficient methods of operation.

I. A QUICK LOOK AT THE INDUSTRY PATTERN

We have been studying various Quality Control systems in use throughout the country in an effort to determine if there is any basic pattern to aircraft industrial Quality Control systems.

A review of a recent Aircraft Industries Association survey of over 20 companies showed a comparison of total Quality Control personnel to the direct labor serviced which indicated that 9 of the companies had a Quality Control force of between 6-1/2 and 9-1/2 per cent of their direct labor force, while the remaining 11 companies ranged between 9-1/2 and 25 per cent, as indicated in Figure 1.

This first stage of the evaluation failed to indicate any basic pattern of similarity between the 20-odd companies studied.

In a comparison of inspectors to direct labor workers of a similar type such as in sheet metal fabrication areas, we found that 10 companies were operating with an inspection force within a range of 4 to 6 per cent, 1 company was operating with an inspection force just under 3 per cent and 10 companies were operating in a range of between 6 and 21 per cent, as indicated in Figure 2.

In a study of the distribution of the Quality Control organization between staff and inspection personnel, it was indicated that 10 of the companies had between 15 and 22 per cent of their Quality Control force in staff operations, 7 of the companies had between 10 and 15 per cent of their Quality Control force in staff operations and 6 of the companies had less than 10 per cent of their Quality Control force in staff operations, as indicated in Figure 3.

This comparison of these three general categories failed to pro-

vide indications of any basic pattern of similarity between the companies studied. However, it was interesting to note that the 3 companies which had the lowest Quality Control forces ranging between 6-1/2 and 7-1/2 per cent of direct labor also had the highest per cent of their total Quality Control forces (between 20 and 22 per cent) in staff operations.

I am not inclined to consider this an industry pattern; however, I do feel that with the implementation of statistical control techniques which are designed to provide a greater degree of inherent product quality, the Quality Control Departments will have a large percentage of their organization in staff positions and fewer people actually making physical measurements.

II. A LONG RANGE TREND OF COMPARATIVE QUALITY CONTROL LABOR COSTS VERSUS RESULTS

A study of methods or means of reducing Quality Control labor costs is of little value unless one simultaneously studies the results of the system in fulfilling the company's objectives as to product quality and customer satisfaction.

The charts in Figure 4 were made recording 3-year cycles showing Quality Control labor costs as a per cent of direct labor costs, the average number of inspection squawks per unit, the average number of customer squawks per unit and the average warranty adjustment claim value per unit.

These charts reflect several things of interest to those working with Quality Control systems and budgets.

It is noted that there was an increase of about 25 per cent in Quality Control labor costs between Cycle 1 and Cycle 2 of these charts, but the inspection squawk rate was reduced by about 1/2 and the customer squawk rate was reduced by about 1/2, yet the warranty adjustment rate went up slightly. I do not know the actual reason for the indicated results as shown on these charts for 1951, other than the fact that there was a rather large company expansion.

The reason for showing these charts is to indicate that the massing of inspection manpower is not necessarily the answer to effective Quality Control.

For example, these charts indicate that in 1954, the company operated at about a 20 per cent lower cost for Quality Control labor as compared to the direct labor dollar than it did in 1948. It is also noted that the average number of customer squawks per unit was about 60 per cent less in 1954 than it was in 1948, and the warranty adjustment claims value was reduced by over 50 per cent in the same period.

In the period between 1951 and 1954, there was a major system and technique change in our Quality Control operation, and a budget system was incorporated into our company operations.

III. STIMULATION OF COMPETITIVE SPIRIT AND CENTRAL GUIDANCE

With the implementation of a budget system at Beech Aircraft

Corporation in 1953, by Mr. Frank E. Hedrick, Vice President and Coordinator, an attitude of competitive spirit was created through the establishment of a relative position between departments as to their budget standing.

For example, in Figure 5, Item 6, is shown the relative budget position in relation to the 24 departments under budget.

Item 4, Column 3, indicates the accumulative dollar value of above or below budget position of the department.

I feel that weekly budget status reports are of major importance in stimulating departmental interest in developing greater efficiency to help make your company more competitive and to better secure the future for both yourself and the company.

IV. THE BUDGET PATTERN FOR QUALITY CONTROL

I have found no magic formula for arriving at a Quality Control budget and, in most cases that I am aware of, the budget is a negotiated annual sum or a certain per cent of the direct labor used with adjustments for varying degrees of outside manufacture of subassemblies or components.

Annual budget adjustments may be made to help make the company more competitive if experience indicates that such adjustments can be made without endangering the products' quality standard.

V. SELECTIVE DISTRIBUTION OF EFFORT

The distribution of the Quality Control effort within the organization basically becomes a matter of evaluating the effectiveness of each control area and the manpower used to provide the necessary degree of control.

For example, at Beech we tabulate on IBM punch cards all of the rejections, our inspection squawks, the departmental responsibility and our customer squawks. A review of these records indicates on a weekly basis the effectiveness of control within each department.

Through the study of this area control record, we can determine the advisability or desirability of increasing or decreasing the inspection personnel in the area, or perhaps the advisability of revising the control technique in a particular area.

VI. DEVELOPMENT OF TECHNIQUES TO INCREASE EFFICIENCY

The budget principle adds a stimulus for directing one's attention towards the exploration of areas of increasing efficiency.

For example, when a basic system appears to have been perfected to the point of near maximum efficiency while maintaining the desired product quality standards, the only potential increase in economy will be from the development of new techniques.

In this area, we have found at Beech that our data from IBM punch card recording of our rejections and squawks gives us a good starting

point for determining the most fertile area for technique improvement.

We have developed a system of Engineering Classification of Characteristics in which Engineering spells out in the engineering data the degree of control necessary to fulfill the designed function of the part. This technique has reduced the direct inspection workload by as high as 44 per cent on some projects, and over a year's experience has proven to have caused no adverse effects on product reliability and customer satisfaction. Actually the customer satisfaction and product reliability have increased; however, we cannot say that this is due to this new technique. We can say that the competitive budget system was a stimulus that encouraged the development of this time-saving technique.

VII. PLANNING FOR THE FUTURE

We are planning for the future at Beech. I do not believe that industry has even approached a realization of either the effectiveness or efficiency that is available as a result of the implementation of modern Quality Control techniques.

We must not be misled by partial evaluations and misconceived economies, for what may be a savings in one area may lead to waste or loss in another area; therefore, it is of paramount importance that all factors are evaluated in determining the cost savings.

For example, at least the following conditions should be compared in analyzing the efficiency of a Quality Control system:

A. Cost of Quality Control manpower
B. Scrap rate
C. Rejection rate
D. Number of incomplete operations turned out by departments
E. Out-of-position rejection rate
F. Squawk rate
G. Customer squawk rate
H. Warranty adjustment rate
I. Effectiveness of corrective action

When your system reflects improvements and savings in each of these areas, I feel you can truthfully say your Quality Control system is gaining efficiency.

I have not yet formed any conclusions as to what can be considered a really effective and efficient system. In the past year, we at Beech have shown indicated improvement in nearly every area mentioned above, and I feel confident that we will show more improvement in the coming year. We feel that improvement is advancement, and we are looking forward to greater improvements in both the effectiveness and economies to be realized through the further implementation of our Quality Control program.

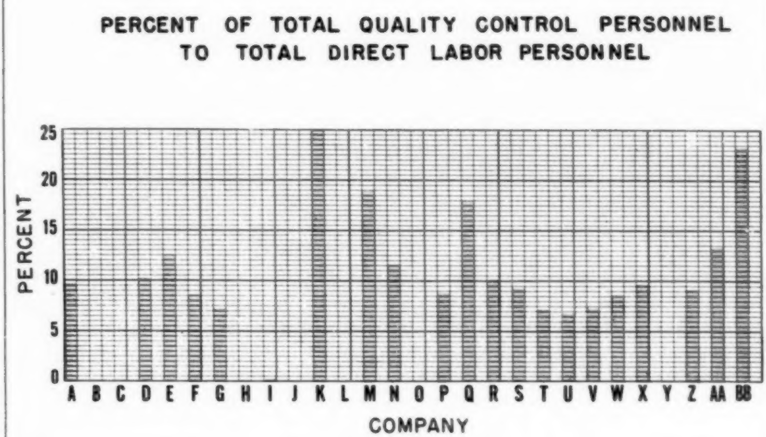

PERCENT OF TOTAL QUALITY CONTROL PERSONNEL
TO TOTAL DIRECT LABOR PERSONNEL
PERCENT
0
5
10
15
20
25
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA BB
COMPANY

FIG. 1

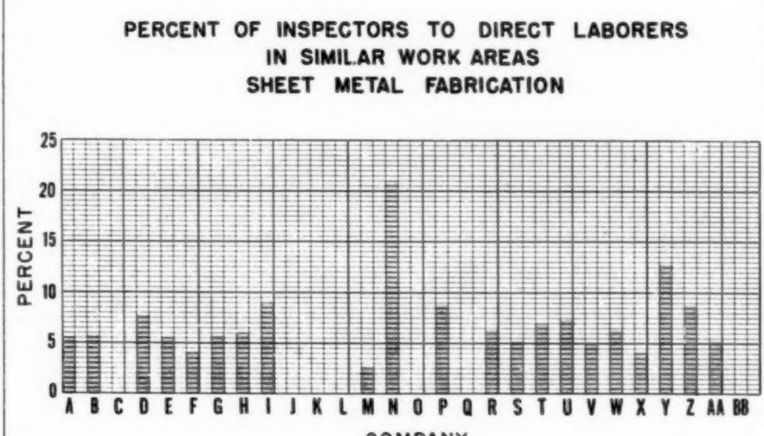

PERCENT OF INSPECTORS TO DIRECT LABORERS
IN SIMILAR WORK AREAS
SHEET METAL FABRICATION
PERCENT
0
5
10
15
20
25
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA BB
COMPANY

FIG. 2

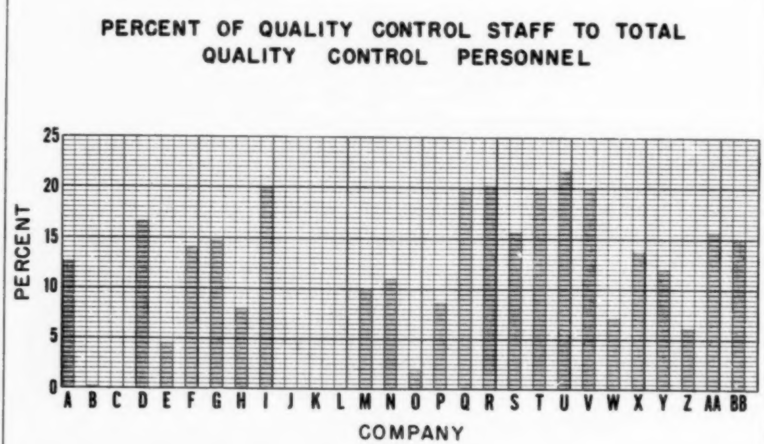

PERCENT OF QUALITY CONTROL STAFF TO TOTAL QUALITY CONTROL PERSONNEL
PERCENT
0
5
10
15
20
25
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA BB
COMPANY

FIG. 3

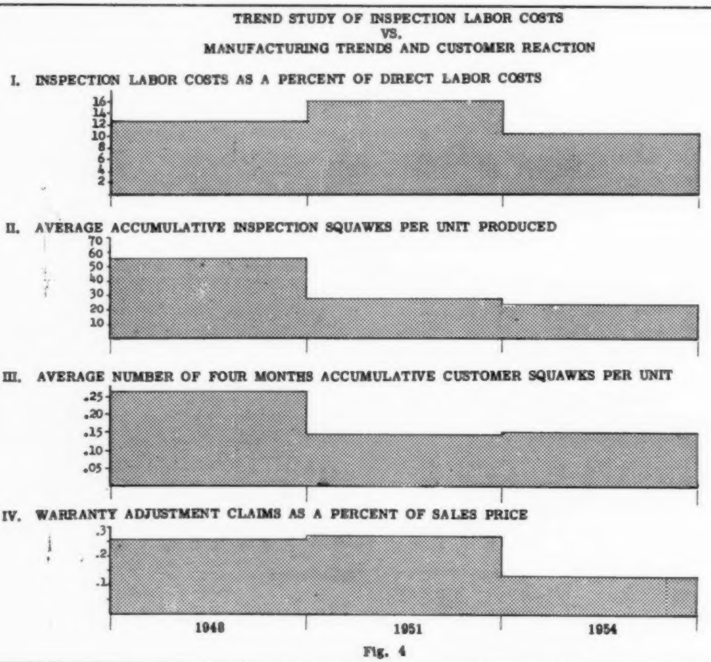

TREND STUDY OF INSPECTION LABOR COSTS
VS.
MANUFACTURING TRENDS AND CUSTOMER REACTION
I. INSPECTION LABOR COSTS AS A PERCENT OF DIRECT LABOR COSTS
16
14
12
10
8
6
4
2
II. AVERAGE ACCUMULATIVE INSPECTION SQUAWKS PER UNIT PRODUCED
70
60
50
40
30
20
10
III. AVERAGE NUMBER OF FOUR MONTHS ACCUMULATIVE CUSTOMER SQUAWKS PER UNIT
.25
.20
.15
.10
.05
IV. WARRANTY ADJUSTMENT CLAIMS AS A PERCENT OF SALES PRICE
.3
.2
.1
1948
1951
1954

Fig. 4

STIMULATION OF COMPETITIVE SPIRIT AND CENTRAL GUIDANCE

Department Head: P. E. Allen
Division: Quality Control
Department No.(s): 78, 178, 278, 478

Date: January 26, 1955

		This Week	Last Week	16 Weeks Cumulative-Fiscal 1955 To Date
1.	Direct Labor Base	$ 227,434	$ 224,340	-
2.	Allowable Indirect Labor Budget (11.5 % of Direct Labor Base)	$ 26,155	$ 25,799	$ 404,884
3.	Indirect Payroll Dollars Expended	$ 23,057	$ 22,998	$ 360,986
4.	Over (Under) Budget - Dollars	$ (3,098)	$ (2,801)	$ (43,898)
5.	Over (Under) Budget - Percent	(11.84) %	(10.86) %	(10.84) %
6.	Relative Budget Standing (See Note II Below)	# 4	# 6	# 6
7.	Actual Percentage to Direct Labor Base	10.14 %	10.25 %	-
8.	Equivalent Personnel (Net)			
	Regular Payroll (See Note III Below)	244	243	-
	Special Payroll (Actual)	19	19	-
		263	262	

RWFisher

R. W. Fisher - Budget Control

cc: Frank E. Hedrick
J. P. Gaty

Fig. 5

THE SAMPLING OF BULK MATERIALS IN THE STEEL INDUSTRY

W. M. Bertholf

There would be little point in merely tabulating the specifications for sampling and analysis of ore, coal, coke, stone, etc. There would be even less in attempting to describe the various methods in use, which vary all the way from the taking of grab samples to quite elaborate mechanical sampling systems. The first is fairly well covered by the handbook "Sampling and Analysis of Coal, Coke and By-Products" (Methods of the Chemists of the United States Steel Corporation) (1) and a paper in Industrial and Engineering Chemistry describing U. S. Steel's methods of sampling iron ore. There are, of course, ASTM Standards covering this subject. The second is pretty well covered by Hassialis in Section 19 of Taggart's "Handbook of Mineral Dressing." (2)

The writer's experience with raw materials for steel making extends back to 1927, when he first came in contact with methods of evaluating iron ore and limestone. The last half of this period has been spent at a coke plant which has a central washing plant for all coals used in the production of metallurgical coke for an integrated steel plant.

If any one thing has been evident for most of that time it is that raw materials are constantly changing, either because of or in spite of our intentions.

We may find that a certain ore, stone or coal is undesirable, either because of its actual or relative quality, and take the necessary steps to replace it with another. Usually, no sooner is this done than it appears that another change is desirable.

One would be happy to report that this is all done with a minimum of sampling expense and that no particular difficulties are encountered in finding acceptable replacements for the undesirable materials. Such is not the case. Constant checking of the quality of raw materials, operating conditions and product is essential. There are times when it seems that the necessary data are as bulky as the materials.

Is this because sampling of these materials is essentially inefficient? We think not.

Almost anyone who has tried to keep up with conditions knows that there are long-term trends in almost every set of time-series data. For example, as shown in Table I (below) there was a progressive decrease in iron content and a corresponding increase in silica content of iron ore shipped from the Lake Superior district in the years 1939-1951. (3)

Table I. Data on Iron Ore

Year	Million Tons	% Iron	% Silica	Year	Tons	% Iron	% Silica
1939	45	51.75	8.27	1946	59	51.32	8.83
1940	63	52.09	8.00	1947	77	50.91	9.09
1941	80	51.83	8.18	1948	83	50.49	9.30
1942	92	51.65	8.21	1949	69	50.39	9.72
1943	85	51.58	8.32	1950	79	50.38	9.85
1944	81	51.72	8.42	1951	93	50.25	9.87
1945	75	51.69	8.52				
				All		51.20	8.84

It is immediately apparent that the quality of iron ore from the Lake Superior district has gradually changed. The samplings on which this conclusion is based simply confirm the obvious conclusion which must be drawn from detailed examination of assay maps of the district.

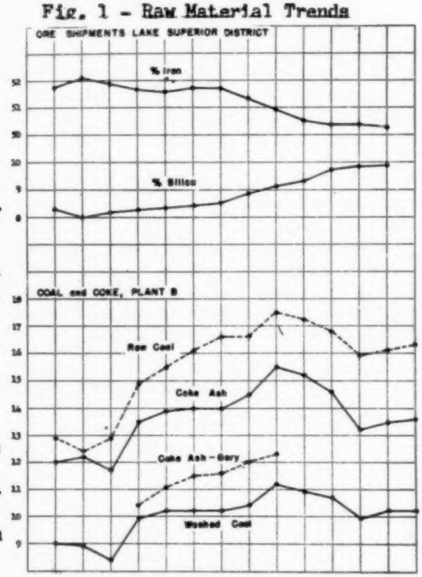
Much the same thing has happened to the coals used for the manufacture of metallurgical coke. An example is given in Table II, the data for which is taken from company files. Confirmation of the opinion that this was not an isolated instance is given by the inclusion of data from an "Eastern" plant for the period 1942-47.

Table II. Coal and Coke Data.

Year	Coal, % Ash Raw	Coal, % Ash Washed	Coke, % Ash	Eastern Coke Ash	Year	Coal, % Ash Raw	Coal, % Ash Washed	Coke, % Ash	Eastern Coke Ash
1939	12.9	9.0	12.0		1946	16.6	10.4	14.5	12.0
1940	12.4	8.9	12.2		1947	17.5	11.2	15.5	12.3
1941	12.9	8.4	11.7		1948	17.2	10.9	15.2	
1942	14.9	9.9	13.5	10.4	1949	16.8	10.7	14.6	
1943	15.5	10.2	13.9	11.0	1950	15.9	9.9	13.2	
1944	16.1	10.2	14.0	11.5	1951	16.1	10.2	13.5	
1945	16.6	10.2	14.0	11.6	1952	16.3	10.2	13.6	

Fig. 1 shows the time trends for ore, coal and coke. It is obvious that one of two things happened (assuming standard methods of taking and preparing the samples), either the materials changed or there was a progressive laboratory error. In view of the fact that blast furnace operating data supports the view that the material changed, it is not likely that laboratory (or operator) error had much to do with the over-all change.

Fig. 1 - Raw Material Trends



In Fig. 2, following page, we show the scatter diagrams and free-hand regression lines for the various pairings of the data. As might be expected, we have no difficulty in using linear regressions for Iron vs. Silica and Coke Ash vs. Washed Coal Ash. However, in the cases of Washed Coal Ash vs. Raw Coal Ash and Coke Ash vs. Raw Coal Ash we must use curved regression lines. In general, it is not good economy to wash a low-ash coal as "well" as one washes a high ash coal, measuring by the reduction in ash content.

Over-all, the yearly data is generally in accordance with facts How would we fare on short-term data? Here we may run into difficulties, in particular if we are trying to check some specific item How universal this condition is can only be conjectured.

Somewhat later in the paper we shall show that there are "cycles within cycles" and that the likelihood of obtaining good checks on small batches of "uncontrolled" materials is not great.

Another complication of the situation is attributable to changes in reagents or apparatus, to say nothing of the individuals involved. If two or more laboratories are involved the situation may get rather messy.

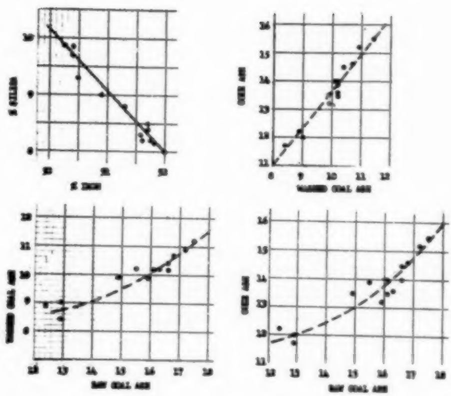


Fig. 2. Regression of Related Characteristics of Raw Materials.

The salient features of several tests on iron ore sampling, in which a comparison of two independent samplings of what was presumably the same material was made, are given below. In Table III the samplings are indicated as "A" and "B", with the "A" sample used as the "control", since it is the sample of "inbound" material.

Table III. Comparison of Repeat Samplings, Iron Ore.

Case	Tons	%Iron, "A"	% Iron, "B"	Nature of "B" Samples
1	26,000	52.23	52.75	30 @ 20 lb minimum, stopped belt
2	26,000	55.25	55.20	100 @ 20 lb minimum, stopped belt
3	30,000	53.5	55.3	400 @ 175 lb, belt running
4	38,000	45.2	47.0*	30 @ 24 hourly increments, size not specified
5	1.5×10^6	59.5	59.9	Total weight 200,000 lb. No data on number of analyses.

* The "B" sample may have been contaminated with ore different from that in the "A" sample. It is stated that there is an 85% agreement in source.

It is immediately apparent that the precision of these comparisons is not always of the order of precision of the yearly averages of Table I. It is a bit risky to hold post-mortems on data which are not adequately identified as to source, but the situation appears to be about as follows:

In cases 1 and 2 the samples were taken in a very short period of time, probably not to exceed 4 days. Since all increments were taken from a stopped belt and represent a complete cross section of the belt load, the conditions for securing a good sample were about as well fulfilled as one could expect, and the results show it.

In case 3 the volume and number of samples could have counterbalanced their intended purpose by bringing in operator and technique effects which would not normally be present.

In case 4 it is obvious from the description originally given that not only was there a chance for "contamination" of the "B" sample but that two different laboratories were involved. It is a bit unusual for the user to report a higher quality material than the shipper.

In case 5 the description of the sampling is so vague that no conclusions can be drawn. Ordinarily one would expect that extensive sampling of such high-grade material would be more precise than this.

A rather striking example of the differences that can happen when a group of laboratories are given a large number of identical samples has been reported previously by the writer (4). The following averages were obtained in a balanced experiment on the determination of ash in coal. Almost 200 samples were split into eighths by a standard procedure and each laboratory was given a split from each sample, with as nearly as possible equal numbers of splits 1, 2, etc.

Table IV. Comparison of Splits and Laboratories, Coal Ash.

Split No.	Av'g. Ash	Split No.	Av'g. Ash	Lab.	Av'g Ash
1	10.20	5	10.23	A	10.3
2	10.25	6	10.19	B	10.3
3	10.22	7	10.21	C	10.6
4	10.25	8	10.20	D	10.1

Obviously the difference between the splits could not have amounted to more than a few hundredths of one per cent, but the differences between laboratories are significantly different. However, it should be noted that for "internal use" the results of any of the participating laboratories would be fairly acceptable. The laboratory differences were almost constant throughout the experiment.

The problem of inter-laboratory standardization is receiving serious consideration in several professional groups and large organizations at the present time. Satisfactory solutions are not yet in sight. It is not likely that everyone will be able to scrap what appears to be useful equipment and start over again, especially if they are getting by very nicely on their present procedures--which is another way of saying that the primary purpose of sampling is to provide useful data. Since all the laboratories participating in the experiment just considered were able to distinguish between low-ash and high-ash coals a change which merely brought their averages closer together would not necessarily be advantageous.

Having shown that there are more or less well-grounded reasons for questioning the absolute accuracy of averages of analyses on samples which are heterogeneous as to source, methods or sampling and analysis and time, what do we have left? In our opinion we are not necessarily in the middle of the ocean on a dark night. Many sets of comparisons between raw material statistics and operating data show definite relationships, not necessarily linear, which can frequently be used to good advantage in "controlling" the process.

One relationship which does not appear to have been considered at length in published data (and in some cases the reasons are quite understandable) is that between the variation in raw materials and quality of finished product. We have seen many correlation or regression analyses

which show the effect of different levels of raw material quality on the operation of the blast furnace and open hearth. The data relating variations appear to be scanty.

Several examples will now be considered, the data being taken from our own files and from the report "Coke Evaluation Project," published by the American Iron and Steel Institute and American Coke and Coal Chemicals Institute as Contribution to the Metallurgy of Steel, No. 43. The limitations of space and time forbid more than an attempt to scratch the surface and reveal some of the pure gold beneath it.

Our first subject is the effect of coke uniformity on blast furnace production and quality of product, to be followed by consideration of the effect of uniformity of ore. A short outline of the methods used in one plant to secure increased uniformity of coke is also given.

We shall consider three blast furnace plants, A, B and C, for which we have coke and operating data covering relatively short periods in considerable detail. It will be shown that the response of the blast furnace is relatively rapid, and that it should not be necessary to run tests over extremely long periods to get an adequate idea of the effect of major changes in quality, particularly if they are sustained.

The writer's interest in this matter is along the line of determining how much fluctuation in coke quality a blast furnace can stand if the changes are "stochastic", which is understood to mean that the changes are not permanent. The correct answer to this question will permit a realistic approach to the problems of proportioning and mixing coals to be used for the manufacture of metallurgical coke.

Figs. 3, 4, 5 and 6 show certain selected coke and blast furnace statistics in decreasing order of "general wildness" The data for Plant A is not typical of normal operations—there was a method in the apparent madness. The first part of the A data covers a period in which an unwashed coal was used for the manufacture of furnace coke with no apparent (or stated) intent to blend out the rather wild changes that are almost inevitable in such cases. The second part of the A data covers a period in which a blended washed coal was used to produce the coke. Standard practice at this plant would have been somewhere between the two extremes.

Case B is recent operating data from our own plant.

Case C is the best we could find.

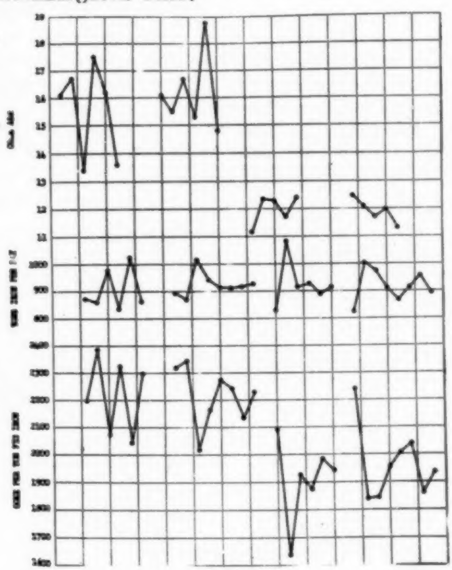


Fig. 3. Coke and Operating Data, Case A

The upper panels of Figs. 3, 4 and 5 show the ash in the coke in chronological order. At Plant A daily samples were composited over the 3 operating shifts. At Plant B coke is sampled on the first and second shifts, with 3 or 4 determinations averaged. Plant C reports the analysis of a composite of hourly samples for each shift. It is obvious that the true variability of Coke A from hour to hour could well have been greated than these data indicate. Coke B appears to have short-time trends within "natural" limits. Coke C has almost the same range of variation as B, but the trend is not so pronounced.

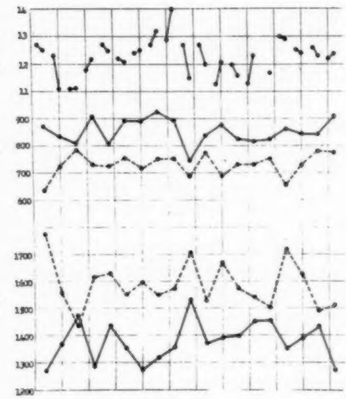


Fig. 4. Case B

The middle panels indicate the tons of iron produced per day. In all cases there are 5 casts per day. Tonnage for each individual cast was not available, hence our data are not quite as sensitive as could be desired. While there are faint traces of trend, nothing conclusive is established. It might be well to note that at Plant A the "wind" was cut when the change from raw coal to washed coal was made, apparently with the idea of maintaining a constant tonnage rate for the entire test period.

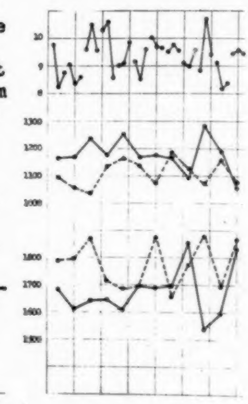


Fig. 5. Case C

The lower panels show the reported pounds of coke per ton of pig iron. The expected inverse correlation is quite pronounced, in general. Only in Case A does the influence of the level of coke ash on coke per ton of pig become conclusively evident. In this case, a drop of about 4 per cent in coke ash decreased the coke usage about 300 lb per ton of iron.

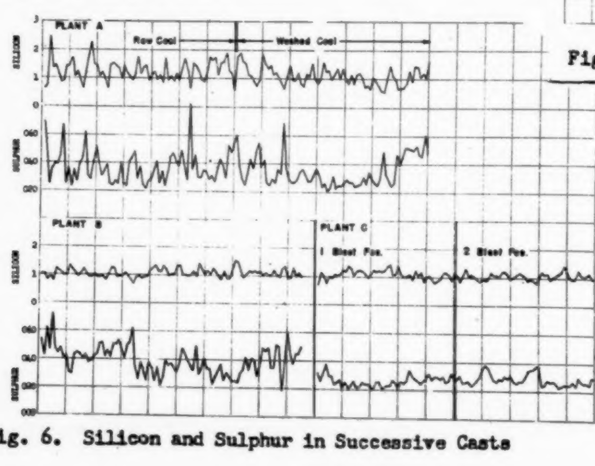


Fig. 6. Silicon and Sulphur in Successive Casts

Fig. 6 is, in our opinion, the most interesting display of data. The silicon and sulphur for each cast of pig iron for the periods in question are detailed, as of the time of cast. As in the preceeding Figs., there is a definite time lag between the reported changes in coke and the appearance of the expected effect at the blast furnace. This is roughly two days in most instances.

The upper panel of Fig. 6 divides equally between the raw coal and washed coal periods on the calendar basis. The time-lag in furnace response is very evident. There is a very evident decrease in variation in iron analysis starting four days after the change in coke (at the ovens) and about two days after the new coke hit the furnaces. The shift in average sulphur at the end of the period is apparently intentional.

The lower left panel of Fig. 6 shows the iron analyses for Case B, data being for the furnace whose tonnages and coke rate are shown as solid lines in Fig. 4. The shift in average sulphur which starts on the 8th day is due primarily to juggling the stone to keep up with a change in ore.

The lower right panel of Fig. 6 shows the iron analyses for the two furnaces of Plant C which operated on the coke for which ash data is shown in Fig. 5. Only the first 10 days of the 14 are shown, but the remaining 4 days were about what one would expect.

It is hoped that we have succeeded in demonstrating that the behavior of a blast furnace is quite likely to reflect the bahavior pattern of the raw materials sent to it. In all these cases the ore was presumed to be relatively constant from day to day. Coke was the major variable in these cases.

Since iron ore is also known to affect the operation of a blast furnace, let us turn our attention to a specific case. Williams (5) has reported on the effectiveness of our ore bedding system. We reproduce one of his tables below, with additions from current practice.

Table V. Blast Furnace Operating Data: Before and After Bedding Ore.

Stage of Preparation of Material:-	1	2	2	3
	"E" Furnace		"F" Furnace	
Blast Furnace Operating Variables	Sept-Oct 1940	Jun-Jul 1945	Nov. 1948	Jan. 1955
Daily Average Iron Variation,				
per cent Silicon	0.57	0.25	0.35	0.11
per cent Sulphur	0.019	0.013	0.019	0.005
Daily Average Slag Variation,				
Silica plus Alumina	1.24	0.75	0.93	0.96
Lime plus Magnesia	1.07	1.26	1.11	1.00
Total Number of Burden Changes				
In metallic mix	12	9	0	0
In stone	32	23	15	8
In coke, or extra coke	151	12	21	4

1) No bedding of ore, no control of coke
2) Ore bedded, no control of coke
3) Ore is bedded, coke is semi-controlled.

The last column of Table V presents new data which, in our opinion indicates rather conclusively that the improved control of raw materials has definitely resulted in a more uniform product. The only section of the above table which does not indicate a definite improvement is that dealing with slag composition.

Williams' original data showed the conversion of a bimodal distribution of raw ore (from two sources) to a quasi-normal distribution. We have taken data from a recent period to show the presence of trends in quality in the ores considered singly. Fig. 7 shows that neither ore A nor ore B is "regularly the same"--the data being daily averages of the ore sent to beds. If it were necessary to use these ores in the proportions in which they are received (by rail, daily throughout the year) even the most scrupulous sampling and analysis would not make it possible to operate the furnaces smoothly. Bedding gives us a chance to iron out the fluctuations in proportion and quality with a minimum of railroad car detention time. Further, it assures the furnace operator that he will have a uniform mix for several weeks at a time.

That this is actually the case is shown by the control charts of Figs. 8 and 9 prepared from data on successive samples of bedded ore, for a different period. The control charts are for Iron content only, but very similar figures are obtained for the other components.

Note that there are no indications of trend in the prepared ore. It is evident that the installation of the ore preparation plant brought the ore situation pretty well under control.

We now turn our attention to the coal and coke situation.

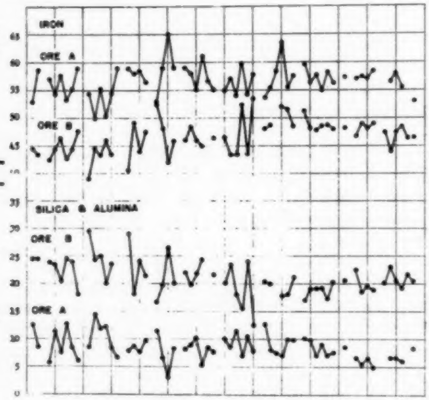


Fig. 7. Fluctuations in Incoming Ore

For many years it was our practice to take as much coal as we could get from our own mines and to purchase any additional coal which might be required to meet the operating schedule of the blast furnaces. This resulted not only in wide fluctuations in the percentage of a particular coal in the mixture but, due to the difference in the character of the coals available at different times, resulted in the fluctuation of coal and coke ash.

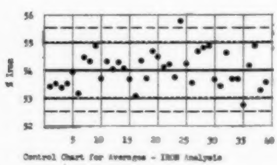


Fig. 8

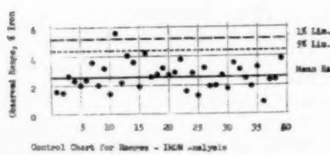


Fig. 9

With the opening of a new company mine, it has recently been possible to eliminate the purchase of outside coals. With a reduction in the number of coals being handled we have been able to proportion the remaining coals more or less scientifically. The over-all results are shown in Figs. 10 and 11, which cover the first half of December in 1951 and 1954. These are neither the worst nor the best examples we could find. In our opinion they are quite typical of the periods under consideration.

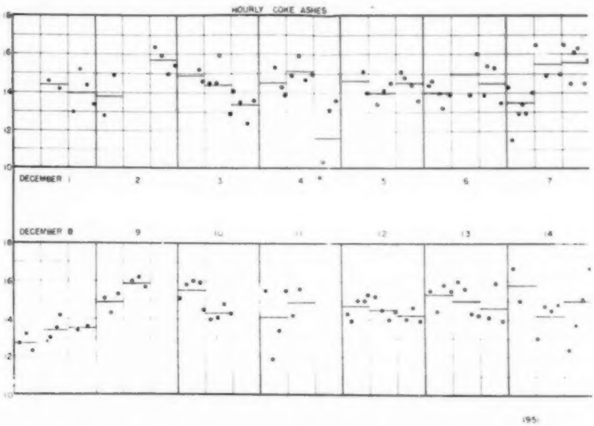


Fig. 10. "Hourly" Coke Ashes, Dec. 1-14, 1951

From Fig. 10, it is evident that in 1951 the average coke ash was between 14 and 15 (for this period) with shift averages ranging from just under 13 to as high as 16. The evidence of trend in the data precludes blaming this on poor sampling. It was unquestionably a case of uncontrolled material. The 10 pct ash coke of the last shift on the 4th is a "sport". On occasion such coke finds its way to the blast furnaces. It is 1 x 3 in. coke from a special wash for foundry coke.

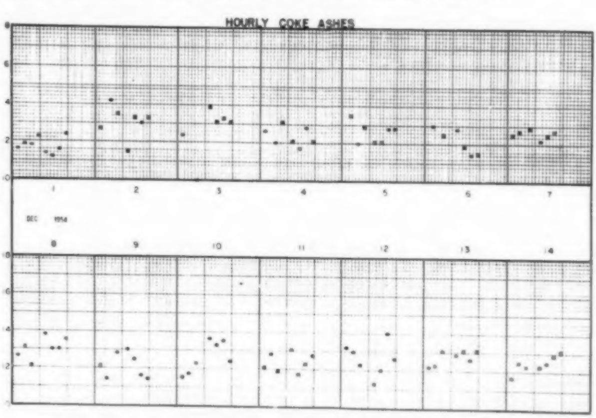


Fig. 11. "Hourly" Coke Ashes, Dec. 1-14, 1954

As shown in Fig. 11, the current average ash is between 12 and 13pct. Much of the trend has been eliminated, or perhaps it would be more accurate to say that the extremes have been pretty well eliminated, or moderated.

That there is a legitimate reason for the apparent reduction in variation in coke ash from 1951 to 1954 is very plainly seen when one considers the reduction in complexity of the coal mixtures used, and the equipment available for proportioning the washery feed.

As shown in Table VI, below, there were, in effect, 15 different coals used in December 1951, divided into 4 categories which were sufficiently different to require that they be used in definite proportions at any given time. With only 5 mixing bins, this required that the 11 heavy coking coals go through 2 bins--in whatever proportions were available at the time. One could hardly expect the mixture to be more than vaguely similar from hour to hour or day to day.

By 1954 we had been able to eliminate the non-coking coal component and the "miscellaneous" heavy coking coals. We now have 5 heavy coking coals going through 3 bins, which is quite an improvement, but not ideal. The resulting increase in uniformity of coke ash is in line with that indicated by a probability analysis of the two situations (using much more data than is given here).

Table VI. Comparison of Coal Situation, 1951 vs 1954.

	December 1951	December 1954
HEAVY COKING COALS:		
Frederick	28.27% a	31.24% a
Morley	9.33	11.52
Allen	1.16	45.11 a
New Mexico	30.76 b	-
Bear Canon	2.28	-
Ludlow	1.65	-
Delagua	5.41	-
From Stockpiles	1.76	-
Sub-total	80.62	87.87
LIGHT COKING COAL	1.40	0.94
NON-COKING	6.73 a	-
LOW VOLATILE	11.25	11.19
	100.00%	100.00%

a) Actually 2 different coals or sources of supply.
b) Actually 3 different coals or sources of supply.

We have examined a considerably body of data relating to the sampling and use of the major raw materials of the steel industry, and it appears fair to conclude that while it is often impossible to completely justify the differences between analyses of presumably the same material reported by different sources the normal "within plant" data covering the quality of raw materials is reasonably close to the truth.

More precisely, there appears to be a good measure of consistency between the variation in quality of raw materials and the quality of the finished product.

If it does not appear to be possible to get "good" samples of raw materials there is more than a bare possibility that the trouble is in the material as much or more than in the sampling. Once the materials are controlled it should not be necessary to employ Superman to get useful information about them.

References.

1) Methods of the Chemists of the United States Steel Corporation for the Sampling and Analysis of Coal, Coke and By-Products, Carnegie Steel Company, Bureau of Technical Instruction, Pittsburgh, Pa.

2) Handbook of Mineral Dressing, Ores and Industrial Minerals, Arthur F. Taggart, John Wiley & Sons, Inc., New York.

3) Minerals Yearbooks, U. S. Bureau of Mines.

4) Tests on the Geary-Jennings Sampler at Cabin Creek, W. M. Bertholf and W. L. Webb, ASTM Symposium on Coal Sampling, to be published in Spring of 1955.

5) Ore Blending, R. R. Williams, Jr., 1949 Proceedings of Blast Furnace, Coke Oven & Raw Materials Conference, AIME, 29 W. 39th St., New York 18, N. Y.

GENERAL SUBJECT MATTER ON THE USE
OF STATISTICAL QUALITY CONTROL
CONFINED TO THE EVALUATION OF MALT
OF DIFFERENT SUPPLIERS

Frank J. Roberts
The Stroh Brewery Company

The production of a uniform and colloidally stable beer is, and has always been, a difficult problem, complicated by many factors - of which the greatest are the variation in raw materials, and the difficulty of determining a correlation between the properties of the brewing materials and the properties of the finished beer. The latter problem is being attacked by many research groups in this country and others, and our knowledge of it is constantly increasing. The widespread employment of pilot plants for this work should prove fruitful, as should the use of modern laboratory developments -- e.g. spectrophotometry, ultracentrifugation, chromatographic techniques, electrophoresis, tracer metals, etc.

The two materials which have the dominating influence on the flavor and stability, as well as foam of the beer, are malt and hops. Hops are particularly important for flavor effect and their employment is controlled by regulating the amount used per brew according to a physical and chemical analysis of the lot. Lots are selected after harvesting, and a whole year's supply is purchased at this time. There is no difficulty in keeping different lots segregated and in varying the quantity used in a brew. This system appears capable of keeping the variation in flavor of beer due to hops at a level unnoticeable by ordinary tasting, although a more thorough understanding of the chemistry of hop constituents and more refined methods for their analysis should further help to reduce the variation.

Malt, however, presents no such fairly satisfactory solution as this. It is received in carload lots and conveyed into large bins. Malt from different suppliers is kept in separate bins, but we do not usually know just which batch of malt is issuing from the bottom of the bins for a particular brew. Even if it were feasible to determine the properties of the malt used in each brew, our limited knowledge of the effect of these properties on the beer would not enable us to vary adequately the brewing process to produce a uniformly stable beer with uniform foam and taste. Therefore, we are forced to keep all the malt received at the brewery as uniform as possible.

This paper will be confined to a discussion of our efforts to improve the uniformity of our beer and its processing through an improvement in the uniformity of the malt from our four suppliers. Naturally, it is a problem which has always received attention, but one which could always stand improvement. It appeared that the principles of Statistical Quality Control would be a valuable help in this problem, and the results we have achieved, with the help of a few elementary statistical techniques, have borne this out.

This discussion will deal with six items of malt analysis - - moisture, extract, diastatic power, alpha-amylase, the ratio of soluble to total protein, and the clarity of malt wort. These are not the only analyses we run on malt, but for the most part, their role in the brew-

ing process and in determining the qualities of the finished beer, are less nebulous than others. Also, by keeping these at a uniform level, we hope that other properties of the malt, for which we have no practical analytical methods and only small knowledge, will also remain uniform.

Since malt is bought on a weight basis, the moisture should be fairly low so that the brewery is not paying money for a lot of water. A high moisture may also subject the malt to the danger of contamination during storage. Moisture also has an influence on the degree of fineness achieved in grinding malt, and therefore should be at a uniform level. Too low a moisture content may cause an excessive amount of flour in the grinding, and lead to a slow run-off in the lautering process. Generally between 4 and 5 per cent is considered suitable.

The extract figure represents the percentage of malt which is soluble in water under standardized conditions of grinding and mashing. Obviously, since malt is purchased on a weight basis, a higher extract in malt will indicate greater economy because more beer will be produced from it. Also, a uniform extract may possibly indicate uniformity in other unmeasured factors.

The three values for diastatic power, alpha-amylase, and soluble to total protein, all represent aspects of malt modification - - a term which is often used, but of which there is at present no clear understanding. It seems that the degree of modification depends on the enzymatic strength of the malt, and also on its susceptibility to enzymatic action. The picture is complicated by many factors; such as barley variety, place of growth, year of growth, malting variations, etc. A distinct change in any one of the three measured items usually will indicate that a change must be made in brew house procedure in order that the beer produced is uniform. Specifically it often means that a different conversion temperature or a different rate of increasing the mash temperature to the conversion temperature must be employed so that the fermentable sugars in the wort, and consequently the percentages of alcohol and extract in the finished beer will be uniform. It may also indicate that a change in the protein rest during mashing will be necessary, perhaps to keep the foam at a uniformly high quality.

Other measures of malt modification, e.g., coarse-grind extract, viscosity of wort, and counts of the percentages of kernels at different stages of growth are also used, but we have not employed these methods as routine procedures in our laboratory.

Clarity of the wort is determined with the use of a nephelometer and the figures represent readings on a nephelos scale. Clear worts will read generally in the range of from 25 to 50. A wort which would be called slightly hazy upon visual examination, would usually give a reading of around 50 to 80 on the instrument, while a definitely hazy wort may be well over 100. A turbid liquid such as Ruh beer will run as high as 500.

The malt received at the brewery within a calendar year corresponds roughly with the barley grown during the preceding growing season. Table I shows the means for the six analytical factors for the two years 1953 and 1954.

The choice of a suitable measure for dispersion of results is not immediately obvious. As barley and malt both undergo changes with age, it is apparent that we cannot expect malt made from new barley to have the same properties as malt made from barley which has been stored for some months. For example, a control chart for diastatic power, such as in Chart I, shows in general, a downward trend through a year's time; although it also shows up and down trends at shorter periods. Such gradual changes with the periodic short-term fluctuations seem to be characteristic of most analytical properties of malt as received at the brewery. Chart II is another example of this, showing that extract values behave similarly.

As a measure of dispersion, we have chosen the standard deviation calculated from a frequency diagram of values for the year's time. These are shown in Table II. The figures in parentheses indicate the standard deviation as calculated from the average range, using subgroups of 3. It is apparent that the latter measure of sigma gives much lower values in most cases, a result of the gradual changes in malt with time, which keeps the range at a low level while the mean gradually shifts.

Practically all items show an improvement in uniformity as indicated by a lower standard deviation in 1954. These improvements were brought about merely by informing each supplier of the figures for 1953 for both himself and the others, and by informing them of any trends away from uniformity as they occurred during the year, with a suggestion that something be done about it. The improvement can probably be attributed largely to the maltsters' developing improved blending techniques, and also to their instituting statistical quality control measures in their own operations.

Control charts were kept on these items. However, we would not be warranted in applying the limits from one year to another year, and hence the charts are used primarily to indicate trends and obvious deviations from uniformity.

We hope sometime in the future, to develop a satisfactory rating scale so that the various analytical values for a shipment of malt can be expressed in just one number, somewhat in the fashion of scoring butter. It would involve first finding an ideal value for each determination and then penalizing in some fashion for deviations from that value. The great difficulty at present, is in knowing just what relative value to place on each factor. Of course, we do have our own opinions, formed largely from practical experience as to which results are the most important, and do apply pressure on a maltster more readily if his malt shows too much variation in such results.

TABLE I

Malt Received at The Stroh Brewery Co. - 1953

	Mean				
Supplier	A	B	C	D	ALL
Moisture	4.59	3.98	4.33	4.26	4.29
Extract (dry basis)	76.80	76.85	76.45	76.63	76.64
Diastatic Power	129.00	122.44	122.00	123.39	124.21
Alpha-amylase	33.05	35.19	33.46	34.63	34.08
Sol./Tot. Protein	38.84	38.35	38.74	38.28	38.55
Clarity	31.66	39.50	32.83	35.40	35.65

Malt Received at The Stroh Brewery Co. - 1954

Moisture	4.70	4.26	4.50	4.54	4.50
Extract (dry basis)	76.69	76.47	76.67	76.81	76.66
Diastatic Power	125.59	125.82	119.08	124.10	123.65
Alpha-amylase	31.57	33.53	32.51	34.96	33.14
Sol./Tot. Protein	37.27	39.93	39.04	38.99	38.81
Clarity	29.82	32.47	30.33	36.33	32.24

TABLE II

Malt Received at The Stroh Brewery Co. - 1953

Standard Deviation

Supplier	A	B	C	D	ALL
Moisture	0.31 (0.17)	0.25 (0.20)	0.21 (0.18)	0.23 (0.26)	0.25 (0.19)
Extract (dry basis)	0.71 (0.50)	0.51 (0.38)	0.52 (0.44)	0.50 (0.32)	0.56 (0.41)
Diastatic Power	6.76 (3.62)	3.94 (3.14)	4.24 (3.38)	5.14 (4.06)	5.02 (3.55)
Alpha-amylase	2.45 (1.23)	1.47 (1.60)	2.04 (1.33)	2.08 (1.25)	2.01 (1.39)
Sol./Tot. Protein	1.36 (0.92)	1.60 (0.89)	1.33 (1.14)	1.25 (0.97)	1.39 (0.98)
Clarity	4.64 (2.71)	7.30 (4.98)	5.28 (3.08)	8.36 (5.71)	6.40 (4.12)

Malt Received at The Stroh Brewery Co. - 1954

Supplier	A	B	C	D	ALL
Moisture	0.21 (0.14)	0.25 (0.21)	0.20 (0.21)	0.21 (0.21)	0.22 (0.19)
Extract (dry basis)	0.53 (0.35)	0.63 (0.34)	0.46 (0.31)	0.39 (0.33)	0.50 (0.33)
Diastatic Power	5.13 (2.35)	5.67 (2.95)	3.89 (3.53)	6.33 (4.32)	5.26 (3.29)
Alpha-amylase	1.51 (0.77)	1.59 (0.92)	1.97 (0.96)	1.21 (1.00)	1.57 (0.91)
Sol./Tot. Protein	1.49 (0.74)	1.34 (1.07)	1.20 (0.73)	1.33 (0.94)	1.34 (0.87)
Clarity	2.48 (1.58)	3.58 (1.95)	3.22 (2.25)	3.90 (2.32)	3.30 (2.03)

(Upper figure represents standard deviations as calculated from frequency distribution. Figures in parentheses represent standard deviation as calculated from R.

CHART I
DIASTATIC POWER--SUPPLIER A

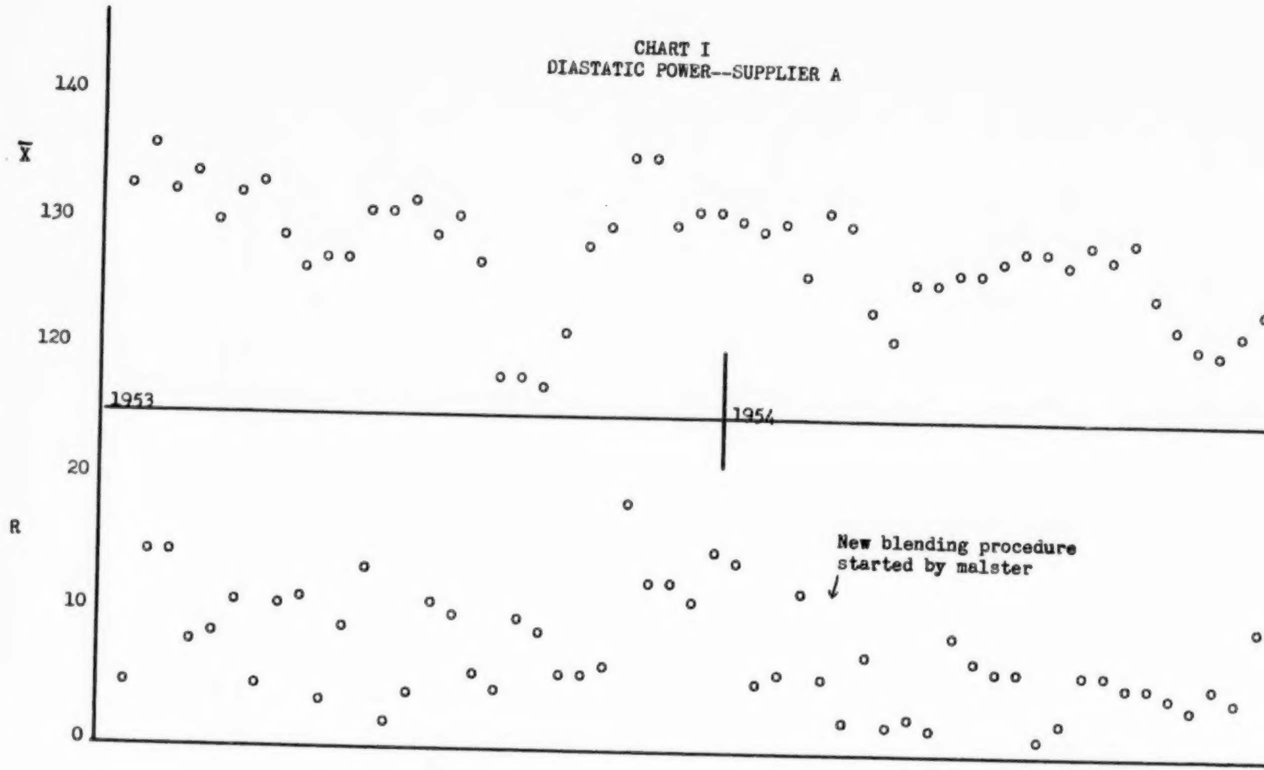
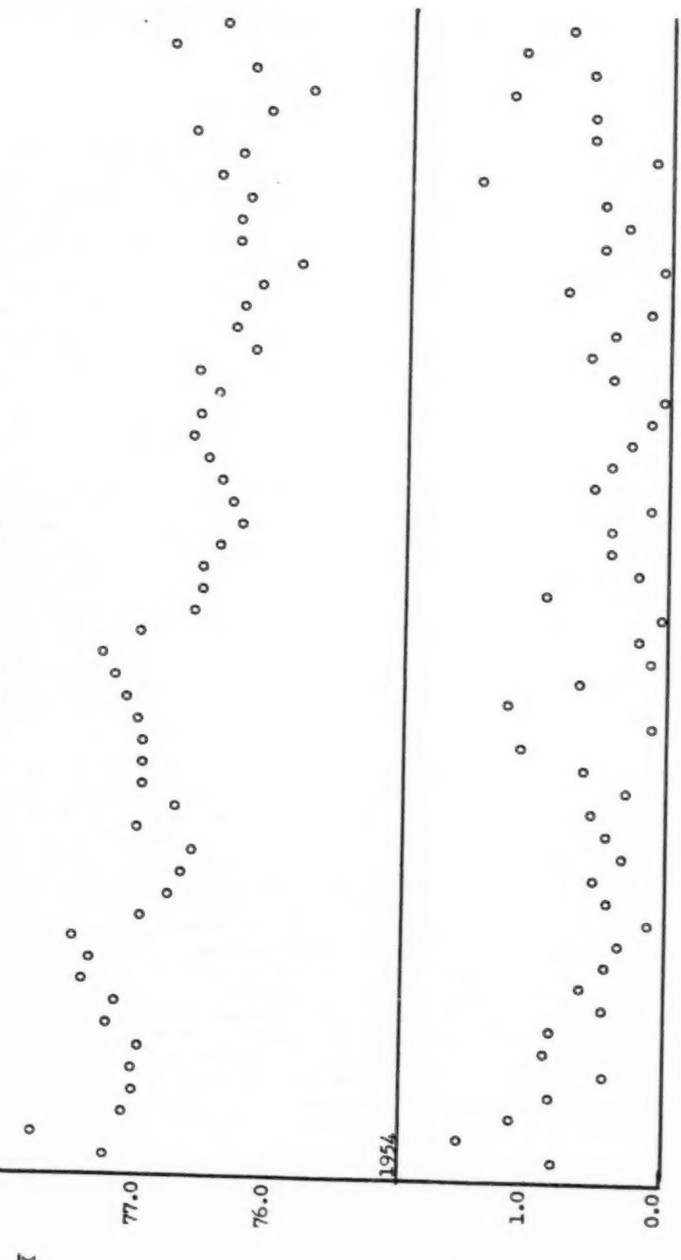


CHART II
EXTRACT - SUPPLIER A



LEGAL ASPECTS OF SAMPLING: RECENT DEVELOPMENTS

Frank R. Kennedy
College of Law
State University of Iowa

While sampling is today a far more familiar tool of the engineer, the business man, and the research scientist than of the lawyer, none of these preceded the lawyer in making use of the representative sample. The law arises out of the fact, says an honored maxim:(1) The generalizations that are the law are grounded on the facts of individual cases. As Alfred North Whitehead has said, "The things directly observed are, almost always, only samples. We want to conclude that the abstract conditions which hold for the samples, also hold for all other entities which, for some reason or other, appear to us to be of the same sort. This process of reasoning from the sample to the whole species is Induction. The theory of Induction is the despair of philosophy--and yet all our activities are based upon it."(2) While the method of the law is traditionally said to be that of deductive logic, the legal process owes much to induction from sample.(3)

It is perhaps not fruitful here to inquire whether sampling theory and practice may be useful in the selection of our representatives in our republican form of government. I assume that constitutional limitations may be incompatible with probability sampling and that, in any event, we seek to select as our representatives those who are better than the mean if not the best among us. It may be worth brief mention, however, that we have become increasingly concerned about the adequacy of the representation achieved by present methods. We have taken constitutional steps to assure that Negroes and women shall be included in the population and inferentially in the representatives that may be selected; and we now are debating whether we should extend like consideration to youngsters old enough to fight.

A recent issue of the Journal of the American Statistical Association describes the mathematical problems involved in apportioning representatives in the lower house of our National Congress.(4) I have recently been engaged in a study of one of our serious political problems in this country--the inequality of representation of urban populations as compared with that of rural constituencies in many of our state legislatures.(5) Our state constitutions frequently embody a compromise of the desideratum of representation proportionalized to population with that of area representation. Constitutional provisions for periodical reapportionment to permit population shifts to be reflected in the legislative representation are quite generally included. Unfortunately a number of state legislatures have ignored their constitutional responsibilities, so that domination of state assemblies by representatives drawn from the rural populations is a common phenomenon today. Since courts cannot force legislators to enact any kind of statute, the need is for a constitutional provision which will accomplish a periodical reapportionment automatically, i.e., without the necessity of legislative action.

It is noteworthy that political processes frequently tend to bring about a representation of particular areas or groups when neither the law nor sound sampling theory would require it or sanction the effort to attain it. An editorial in this morning's edition of the Des Moines Register condemned the Iowa state senate's refusal to confirm the

governor's nominees for the state highway commission because the refusal was based on the governor's alleged indifference to demands for geographical representation.(6) It has been traditional for justices of the Supreme Court of the United States to be selected with a view to achieving "fair" geographical distribution, and it has been suggested that there should be a Jewish member and a Catholic member.(7) Equally good reasons can be advanced for including representatives of other racial and religious groups, women, labor, laymen, ad inf. While surely diversity of origin and interest among the members of a court or almost any other body charged with broad public responsibilities is on the whole to be desired rather than avoided, any substantial emphasis on achieving representation of particular areas, groups, and interests is likely to entail a sacrifice of quality and in the end prove to be a frustrating and hopeless endeavor.

When President Roosevelt in 1937 brought out his so-called "Court-packing plan," a great deal of virtue was pinned to the number nine by the opponents of the plan. The number is of course not fixed by the Constitution, and over the span of the Court's history its number has varied from six to ten. It has remained at nine for the last fifty years, however, and experience has demonstrated its appropriateness, quite without reference to any evidence of its adequacy as a sample. Although there are to be found six-man juries and even one-man juries in this country, the prevailing preference is for a jury of twelve men. The settling on twelve is apparently due to a persistence of an ancient belief in the mystic virtue of the number rather than to a calculated judgment as to the sufficiency of the sample,(8) but again the appropriateness of the size has been vindicated by experience. There have been recent challenges in the courts to the composition of particular juries on the ground that they have not been fairly representative. The assumption underlying such challenges is that a jury of one's peers must constitute a fair cross-section of the community from which it has been chosen.(9) The courts have not accepted the validity of this position. Challenges to the New York "blue-ribbon" jury system for its allegedly systematic exclusion of laborers and women have been twice rejected by the Supreme Court of the United States in recent years.(10) Nevertheless Negroes who have been able to show systematic exclusion of members of their race from the panels from which juries were chosen in their communities have won reversals for denial of equal protection of the laws.(11) And recently the principle was extended to protect an accused of Mexican extraction who established that members of his ethnic class had been systematically excluded from juries in the Texas county wherein he was convicted.(12) Note, however, the limitations on the constitutional doctrine: While the equal protection clause is not limited in its condemnation to discrimination against Negroes, the accused was obliged to show that persons of Mexican descent constituted a separate class in his county. A principal part of the proof was that 14% of the county's population had Mexican or Latin American names. While some persons of Mexican descent qualified for jury duty, none had served in twenty-five years. Chance could not explain the total exclusion from 6000 jurors. The Court rejected the suggestion that it was requiring proportional representation or that the accused should have a person of Mexican descent on the jury trying him. When in another case it appeared that a Negro was deliberately put on a grand jury which indicted another Negro in Texas, the Supreme Court rejected an attack on the indictment predicated on the argument that representation on the grand jury was not proportional.(13) Mr. Justice Murphy thought that such a process of conscious selection violative of the equal protection clause because it

would necessarily result in arbitrary limitation. He thought the Constitution required, not proportional representation, but elimination of the racial factor in the selection process. "This may in a particular instance result in the selection of one, six, twelve or even no Negroes on a jury panel."

After the United States Supreme Court recently reversed the conviction by a Florida jury of a Negro for rape of a white woman,(14) his counsel sought a change of venue for the new trial ordered by the higher court. To substantiate his claim that an impartial jury trial was impossible to obtain in the county where it was scheduled, he sought to introduce the results of a public opinion survey conducted by the Elmo Roper Research and Public Opinion Organization. The tendered data were to be presented by Dr. Julian L. Woodward, a research executive of considerable experience who supervised the survey, and a field representative. 500 white persons and 150 Negroes were selected for interview in the county of the scheduled trial, and for statistical reasons a smaller number of interviews of both whites and Negroes was to be conducted in three other counties. The survey apparently followed familiar patterns for this kind of investigation, quota sampling being employed. Forty-three percent of the 518 people actually polled in the county of trial were convinced of the accused's guilt. The court rejected the entire report and the executive's statements regarding it because they were hearsay on hearsay—neither witness having heard the interviews and no interviewee being in court to be cross-examined. The court acknowledged the propriety of such a survey in respect to consumer attitudes toward trade names and products, but the pollsters' big blunder in 1948 was more impressive. Instead the court accepted the testimony of several witnesses presented by the State, white and colored, all of whom testified that a fair trial could be had. The trial was had, the accused convicted and sentenced to death.(15) The court was justified of course in wondering whether the responses obtained would demonstrate that twelve impartial jurors could not be found and relied on to decide the case on the legal evidence. But the court has been rightly criticized(16) for refusing even to admit the evidence and giving "overwhelming" weight to the testimony of witnesses procured by one side.

While courts have not been unanimous in accepting the results of public opinion surveys, the instances where such data have been accorded judicial consideration are now numerous. Their admissibility in cases of alleged trade-mark infringement and unfair competition to establish consumer understanding or likelihood of consumer confusion with respect to trade symbols is fairly well established.(17) Perhaps the most forthright and impressive precedent supporting admission of survey evidence is found in United States v. 88 Cases ... /of/ "Bireley's Orange Beverage" involving a condemnation by the Food and Drug Administration.(18) The Government introduced surveys of consumer opinion to establish that the product appeared to be better than it was. The surveys were vigorously challenged for utter disregard of the principles of random sampling, for discrepancies between results on differing surveys, and for a half-dozen other reasons, but the Court of Appeals for the Third Circuit held the results to be admissible for whatever weight the trier of the fact might care to give them.

The same result was reached more easily by a New York state court where it was convinced that "probability sampling," "the best method in the sampling art," had been faithfully carried out in a survey designed to establish the public understanding in Nassau County of the words "savings" and "saving" in advertising and publicity for banks.(19) This

court thought the planners, supervisors, and workers (or some of them) should testify, and their work sheets, reports, surveys, and all documents used or prepared during the poll taking as well as those showing the results should be offered in evidence. It may be doubted that so full a presentation is ordinarily called for, but the court's handling of this situation suggests that those engaged in sampling activity that may sometime encounter legal scrutiny would do well to anticipate the possibility of similar judicial eagerness to examine all relevant materials.

The Food and Drug Administration has had a considerable amount of experience in the courts with its sampling procedures. It regularly condemns entire shipments of products on the basis of inspection of samples. The courts have generally shown little sophistication with respect to what proper sampling may require, and the result has generally been to sustain the Government's complaint even though at least theoretically the burden of proof as to the adequacy of the sample is on the Government.(20) It is doubted that any well directed attack against the adequacy of the sampling has ever been marshalled. Counsel opposing the Government must, however, also deal with the question as to how much, if any, noncompliance with the statutory standard may be tolerated. The law prohibits shipment of any article of food or drugs which is adulterated or misbranded.(21) Adulteration on account of the presence of a "filthy, putrid, or decomposed substance" occurs if the product "consists in whole or in part" of such a substance.(22) Judge Learned Hand opined that filth nevertheless had to be present in a substantial degree to satisfy the statute,(23) and a couple of district courts seemed to think that a jury could allow a tolerance for decomposed salmon inasmuch as such a salmon might occasionally get into the canner's product notwithstanding ordinary care.(24) One case sustained a condemnation based on the Government's sampling and examination which indicated that 12 per cent of the shipment was "bad" and 25 per cent stale notwithstanding the fact that two other tests conducted by private parties resulted in a finding of good quality.(25) These two tests corroborated each other and one involved a sample taken by selecting one can from every forty-third case, ten cans in all. The Government's examination involved a substantially larger sample, however, and "was of a more extended and careful character than that given otherwise."

A statutory development less well known perhaps than the federal food and drug legislation is the extensive adoption of legislative commodity standards by the states.(26) The state statutes generally provide for inspection service and deal in various ways with the problem of sampling: Departmental rules dealing with the problem may be authorized;(27) minimum sizes of samples may be fixed;(28) the official methods of sampling prescribed by the Association of Official Agricultural Chemists may be prescribed.(29) While some of these statutes make compliance with the standards compulsory, their principal function is to furnish permissive standards for use in negotiating and drafting contracts and in settling disputes that arise. When parties avail themselves of the opportunity to utilize a permissive statutory standard by agreeing upon government inspection, the buyer is bound by a certificate of the chosen inspector showing conformity.(30) The same result flows of course from a certificate of inspection performed pursuant to an agreement having no statutory basis.(31) In either case evidence may undoubtedly be introduced to show that there was fraud involving the seller in the making of the inspection.(32) While fraud may be inferred from a great disparity between the certified result and that from a second inspection, a certificate may not

be attacked by evidence showing merely that it was mistaken or that the goods actually conformed.(33)

Suppose the statute establishing the standard or the agreement of the parties provides not merely for the selection of the inspector who shall issue the certificate but also prescribes the procedure for the inspection including the manner of his sampling. Can the effect of the inspector's certificate be defeated by a showing that he departed from the prescribed sampling procedure or by a showing that a second inspection based on that procedure reached irreconcilable results? No more definite answer can be given than that the court must seek to ascertain the intent of the parties as disclosed by the words they used and the circumstances deemed relevant to such an inquiry. Clearly the burden should be on the party attacking the certificate to show that any contractual requirement relative to procedure has not been followed.(34) As before noted, gross discrepancies in the results reached on two inspections may be regarded as evidentiary of such fraud or misconduct as to vitiate the certificate based on the first inspection.(35)

Parties to a sale contract may specify in great detail their agreement as to quality, the method of inspection, and the consequences of nonconformity. The Government of the United States exercises its rights as a contracting buyer to particularize in respect to these matters, and it appears to be engaged in extending its degree of control over the production process of its suppliers and their subcontractors and their suppliers. Inspection and test by the Government do not of course relieve the contractor from responsibility regarding defects or nonconformity discovered prior to final acceptance. Final acceptance is conclusive on the Government under standard supply contracts except as to latent defects, fraud, or such gross mistakes as amount to fraud.(36)

Parties may on the other hand purchase and sell without making explicit their intentions as to quality, inspection, and consequences of breach. A vendor is likely, nevertheless, to find that he is chargeable with having warranted the quality to be merchantable--i.e., to be "fair average quality in the trade and within the description" and to "run, within the variations permitted by the agreement, of even kind, quality and quantity within each unit and among all the units involved."(37) The buyer has a right of inspection before payment unless he agrees otherwise, as when delivery is "C.O.D." or a negotiable bill of lading has issued.(38) Even after acceptance, however, he may revoke for nonconformity difficult of discovery before acceptance.(39)

To protect sellers of perishables against unfair and unjustified rejections and demands for allowances by buyers in remote cities, Congress enacted the Perishable Agricultural Commodities Act (40) to facilitate prompt inspection. This opinion is not conclusive, provision being made for arbitration. There may be ultimate appeal to the federal courts, but the findings on conformity are prima facie, though not conclusive, evidence in court. The new Uniform Commercial Code, adopted in Pennsylvania and being proposed elsewhere, gives either the buyer or the seller in the event of a dispute over quality or conformity the right to inspect, test, and sample the goods for the purpose of preserving evidence; or they may agree on a third-party inspection or survey and may agree to make the findings binding in subsequent litigation.(41) Without such an agreement or a statute the results of any inspection have no more force than the finder of the facts deems they deserve. No matter how well designed and executed any plan for acceptance sampling, a purchaser is

entitled to reject what does not comply, to revoke acceptance after reasonably adequate inspection failed to disclose a defect, and to recover damages for any nonconformity of accepted goods.(42) That the manufacturer or seller was without fault in the situation affords no defense. If he would escape the consequences of alleged nonconformity, he must ordinarily meet the buyer's proof on the issue of conformity. Quality control charts, like other test and inspection data, may be persuasive evidence to counteract that adduced by the buyer.(43)

It might be assumed that where no statutes provide facilities and standards for reducing disputes regarding conformity, a good deal could be done by cooperative arrangements of producers and distributors to reduce differences regarding sampling and inspection procedures, standards of quality, acceptable tolerances, and the like. Whenever cooperative activity among business men is suggested or tried, however, the impact of the federal and state antitrust laws must be considered. Such activity may implement a combination which restrains competition. The activity is particularly vulnerable to condemnation under our antitrust laws when it facilitates price uniformity, i.e., the elimination of price competition. The Federal Trade Commission has frequently found that an important preliminary step in the establishment of effective horizontal price-fixing combination among competing producers was the standardization of the products sold by the members of the combination.(44) Cooperative activity designed to eliminate disputes and difficulties and to further economic objectives that are not anti-competitive is legal. Activity having no significant scientific or economic justification other than that of eliminating competition is likely to be illegal. Activities in themselves innocent and justified may be condemned because inseparable from illegal activities. It is likely that the line marking the boundary between legal and illegal conduct under the antitrust laws will remain vague for some time.(45)

Perhaps the most notable development in the judicial use or sanction of sampling has occurred in connection with antitrust proceedings. Because of the complexity of the economic and technological issues of fact and the volume of material that is relevant in this kind of litigation, the necessity for abstraction and for systematic organization, evaluation, and presentation has been appreciated here more than in any other area.(46) The exigencies of the Big Case seem suddenly to have been realized in the last five years.(47) In United States v. United Shoe Machinery Corp.,(48) a "trial of prodigious length," Judge Wyzanski "attempted to shorten the hearings ... by encouraging the use of sampling devices." Sampling was employed in the case to show the defendant's share of the market. The court suggested that the Government take depositions of 45 shoe manufacturers operating 55 factories.

> "The Court arbitrarily selected from a standard directory of shoe manufacturers, the first 15 names that began with the first letter of the alphabet, the first 15 names that began with the eleventh letter of the alphabet, all 8 of the names that began with the twenty-first letter of the alphabet, and the first seven of the names that began with the twenty-second letter of the alphabet. This sample covers 3 per cent of the shoe manufacturers. ... Probably the sample unintentionally over-represented machines used in the cement process, somewhat under-represented those in the Goodyear welt process, and greatly under-represented those used in the stitchdown, Littleway Lockstitch, and some minor processes. But these and any other distortions discussed at this bar, would have the effect of showing

United with a smaller percentage of the aggregate market than a better devised sample. And in criticizing this sample, United has not suggested, much less offered, a preferable sample. If antitrust trials are to be kept manageable, samples must be used, and a sample which is in general reasonable should not be rejected in the absence of a better sample."(50)

Data obtained by sampling were also used in the Aluminum Company of America,(51) the Socony-Vacuum,(52) and the J.I. Case (53) antitrust cases. The Committee on Practice and Procedure in the Trial of Antitrust Cases of the Section of Antitrust Law of the American Bar Association, responding to a challenge to help solve the important procedural problems in antitrust litigation, in its first report approved more extensive use of sampling.(54) It observed that "the relative accuracy of proof by sampling may surpass the oftentimes speculative character of expert opinion evidence based in part on the hypothetical question, which in turn is based upon some partial statistics presented into evidence by other witnesses."(55)

The use of sampling in the courts is increasing.

References

(1) Broom's Legal Maxims 57 (10th ed. 1939).

(2) Whitehead, Science and the Modern World 35 (1926).

(3) Allen, Law in the Making 264 (5th ed. 1951); cf. Levi, Introduction to Legal Reasoning 1, 19 (1949).

(4) Willcox, Methods of Apportioning Seats in the House of Representatives, 49 J. Am. Stat. Ass'n 685 (1954).

(5) Stoyles and Kennedy, The Iowa Senatorial Districts Can Be Reapportioned—Constitutional and Legal Aspects of a Plan, 39 Iowa L. Rev. 557 (1954).

(6) Des Moines Register, Feb. 13, 1955, p. 4-M, cols. 1-2.

(7) Cf. McHargue, Sectional Representation on the Supreme Court (1951) 35 Marq. L. Rev. 13 (1951); Ewing, The Judges of the Supreme Court, 1789-1937 pp. 40-63 (1938).

(8) Lesser, The Historical Development of the Jury System 98 (1894).

(9) Four justices of the Supreme Court apparently once thought this assumption to have a constitutional sanction. See Mr. Justice Murphy's dissent in Fay v. New York, 332 U.S. 261 (1947). The counsel who propounded this view in the last cited case and supported it with elaborate tabulations to show nonrepresentation of the lower economic classes under the New York jury system was Harold R. Medina, who later as judge rejected the same kind of argument when advanced by the Communists on trial under the Smith Act.

(10) Fay v. New York, 332 U.S. 261 (1947); Moore v. New York, 333 U.S. 565 (1948).

(11) See, e.g., Norris v. Alabama, 294 U.S. 587 (1935).

(12) Hernandez v. Texas, 74 S.Ct. 667 (1954).

(13) Akins v. Texas, 325 U.S. 398 (1945).

(14) Shepherd v. Florida, 341 U.S. 50 (1951), invalidating the conviction below because of discrimination against Negroes in the method of selection of the grand jury that indicted two defendants before the Court.

(15) Irvin v. State, 66 So.(2d) 288 (Fla. 1953).

(16) In 29 N.Y.U. L. Rev. 751 (1954) and 52 Mich. L. Rev. 914 (1954).

(17) 29 N.Y.U. L. Rev. 751 (1954).

(18) 187 F.(2d) 967 (3d Cir. 1951).

(19) People v. Franklin Nat. Bank, 200 Misc. 557, 105 N.Y.S.(2d) 81 (Sup.Ct. 1951), rev'd on other grounds, 281 App.Div. 757, 118 N.Y.S.(2d) 210 (2d Dep't 1953), aff'd, 305 N.Y. 453, 113 N.E.(2d) 796

(1953), rev'd, 74 S.Ct. 550 (1954). The claimant unsuccessfully objected in the Bireley case, cited in n. (18), to the turning over of the completed questionnaires to the jury during its retirement. See also United States v. 38 Dozen Bottles, 114 F.Supp. 461 (D.Minn. 1953).

(20) I have discussed the matters referred to in this paragraph at length with citations of authority in Sampling by the Food and Drug Administration, 6 Food Drug Cosmetic L. J. 759 (1951).

(21) 52 Stat. 1042 (1938), 21 U.S.C. § 331 (1950).

(22) See, e.g., 52 Stat. 1046 (1938), 21 U.S.C. § 342 (1950).

(23) United States v. 408 Bushels of Oysters, Decis. of Cts. in Cases Under Fed. Food and Drugs Act 718, 720-1 (D.N.Y. 1916).

(24) United States v. 496 Cases of Canned Salmon, Decis. of Cts. in Cases Under Fed. Food and Drugs Act 1099 (D.Mo. 1924); United States v. 243 Cases of Canned Salmon, id. at 1131 (D.Tex. 1925).

(25) United States v. 430 Cases of Canned Salmon, Decis. of Cts. Under the Fed. Food and Drugs Act 1035 (D.S.C. 1922).

(26) See Note, 31 Col. L. Rev. 872 (1931).

(27) See, e.g., Iowa Code § 189.5 (1954).

(28) See, e.g., Iowa Code § 199.8 (1954) (agricultural seeds); id. § 208.3 (illuminating oil).

(29) See, e.g., Iowa Code § 200.9 (1954) (commercial fertilizers).

(30) Field v. Descalzi, 276 Pa. 230, 120 Atl. 113 (1923). The seller is likewise bound by the certificate if it shows noncompliance. Van Dusen Harrington Co. v. W. F. John & Co., 127 Wash. 426, 221 Pac. 301 (1923). But cf. International & G. N. Ry. v. Diamond Roller Mills, 36 Tex. Civ. App. 590, 82 S.W. 660 (1904)(terminal carrier not bound).

(31) Cf. Tacoma & Eastern Lbr. Co. v. Field, 100 Wash. 79, 170 Pac. 360 (1918).

(32) Hettler Lbr. Co. v. Olds, 221 Fed. 612 (6th Cir. 1915); Tacoma & Eastern Lbr. Co. v. Field, 100 Wash. 79, 170 Pac. 360 (1918).

(33) Federal Grain Co. v. Hayes Grain & Comm'n Co., 161 Ark. 51, 255 S.W. 307 (1923)("gross mistake" of inspector not fraud); Citizens' Independent Mill & E. Co. v. Perkins, 52 Okla. 242, 152 Pac. 443 (1915).

(34) Cf. Comment, 34 Iowa L. Rev. 526 (1949).

(35) See note 32, supra.

(36) See generally Lupton, Government Contracts Simplified 250-8 (1954). For a critical analysis of Government contracting practices in respect to inspection and quality control, see National Security Industrial Ass'n Report to Commission on Organization of the Executive Branch of the Government Regarding Military Procurement 75-81 (1954).

(37) The quotations are from Uniform Commercial Code § 2-314(2)(b) and (d). Although the Code has not yet been extensively adopted, its provision regarding the implied warranty of merchantability purports to be drawn from the developing case law of sales.

(38) Vold, Sales 424 (1931).

(39) Vold, Sales 497 (1931); cf. Uniform Commercial Code § 2-608.

(40) 46 Stat. 531 (1930), as amended, 7 U.S.C. § 499 et seq. (1950).

(41) § 2-515. The cited section and the Act cited in the preceding note are discussed in a Note, 20 U. of Chi. L. Rev. 125 (1952).

(42) Vold, Sales 491 et seq. (1931).

(43) I discuss this matter more fully in some Legal Aspects of Sampling, Industrial Quality Control, Jan. 1951, pp. 24-31.

(44) See, e.g., Crown Mfrs. Ass'n, 45 F.T.C. 89 (1948), aff'd as modified, 176 F.(2d) 974 (4th Cir. 1949).

(45) For related discussion see Sonnett, Is Standardization Legal?, 19 Indust. Standardization 192 (1948); Timberlake, Standardization and Simplification Under the Antitrust Laws, 29 Corn. L. Q. 301 (1944).

(46) See Dession, The Trial of Economic and Technological Issues of Fact: I, 58 Yale L.J. 1019 (1949).

(47) See bibliography included in Report of the Committee on Practice and Procedure in the Trial of Antitrust Cases of the Section of Antitrust Law of the American Bar Association 58-60 (1954).

(48) 110 F. Supp. 295 (D. Mass. 1953).

(49) Id. at 298.

(50) Id. at 305.

(51) United States v. Aluminum Co. of America, 35 F.Supp. 820 (S.D.N.Y. 1940).

(52) United States v. Socony-Vacuum Oil Co., 310 U.S. 150 (1940).

(53) United States v. J. I.Case Co., 101 F.Supp. 856, 868 (D. Minn. 1951), where, however, the court doubted the relevancy or materiality of the data but allowed it in evidence "for what it may be worth."

(54) Its Report was submitted on May 1, 1954. Its discussion of sampling and opinion polls appears at pp. 46-49.

(55) Report at 47, citing United States v. American Tobacco Co., 39 F.Supp. 957, 958 (E.D. Ky. 1941), where expert opinion founded on speculative evidence was excluded.

SIGNIFICANCE TESTS BY RANK METHODS

Frank Wilcoxon
Lederle Laboratories Division
American Cyanamid Company

The idea of using ranks instead of the actual measured values is quite old, and in a footnote to a paper by Kruskal and Wallis (4) it is stated that ranks were used as early as 1778 by Laplace. The earliest systematic rank test is probably the Spearman rank correlation coefficient. Spearman published a paper on "The Proof and Measurement of Association between Two Things", and the Spearman rank coefficient was studied by Student in 1920 (6).

In 1938 Kendall proposed a different rank correlation coefficient (3), based on inversions of order. In 1945 the present author (8) proposed to use rank methods to test whether two samples could have been drawn from the same population. Two rank methods were suggested. One for the case of two groups of replicates, not paired with each other, and another for the case of paired measurements. Such methods have often distinct advantages, since they do not require the assumption of normality of the populations from which the samples are drawn. Very little computation is required in making the test of significance, and they may be used in cases where the original data are in the form of ranks or scores. The following table shows a comparison of the flexural strength of two resin laminates, A and B. There were 10 samples of each kind.

	A	rank		B	rank
	20,500	15.0		19,500	9.0
	19,000	6.5		18,100	3.0
	21,200	19.0		17,000	1.0
	21,800	20.0		17,200	2.0
	20,400	13.0		18,200	4.0
	21,000	17.5		20,400	13.0
	20,700	16.0		18,300	5.0
	20,100	11.0		19,600	10.0
	20,400	13.0		19,200	8.0
	21,000	17.5		19,000	6.5
Av. =	20,610	148.5	Av. =	18,650	61.5

The results have been assigned rank numbers running from 1 to 20, and where ties occur each number has been given the average rank. If these two kinds of laminate were really the same, the expected rank totals for A and B would be one half the sum of the numbers 1 to 20, which is 105. The actual sum obtained under B is only 61.5. It is possible to compute the probability of obtaining such a result by chance if the materials were the same. We must enumerate the different ways of getting all possible totals from the lowest total possible which is 55, up to 61 or 62. The sum of these ways divided by the number of ways of getting all possible totals gives the probability of 61 or less under B. This fraction must then be doubled to give the two-sided probability.

It turns out that there are 30 ways of getting a total of 61 or less. The number of ways of getting all possible totals is given by the number of combinations of 20 objects taken 10 at a time. This value is 184,756. The resulting one-sided probability is 0.000162, and the two-sided probability is 0.000324 or about 3 chances in 10,000. We are quite justified in deciding that these two laminates differ in flexural

strength, since if they did not the chance of obtaining the result given above is very small.

In assigning the ranks to the data it is convenient to write the measurements on small plastic chips which can easily be arranged in order of increasing magnitude and the ranks assigned without making mistakes.

It is also convenient to have probability tables computed in advance, and thus eliminate the need of any calculation except the addition of the ranks. Such tables are available in (7) and (9).

It should be pointed out also that this test is applicable to the case where the number of measurements is not equal in the two groups being compared.

In case the number of measurements is 10 or more it is possible to approximate the true probability by taking advantage of the fact that the distribution of rank totals is almost normal with a standard deviation equal to $\sqrt{\frac{N\bar{T}}{6}}$, where N is the number in each group and $\bar{T}$ is the expected rank total if the two materials were the same. In this case N is 10, while $\bar{T}$ is 105. The deviation from the expected value is 44 for a total of 61 while the standard deviation is $\sqrt{175}$, or 13.229. The ratio 44/13.229 is 3.326. The two-sided probability obtained from tables of the normal probability integral is 0.000881, or about 9 chances in 10,000.

In the next example we have a somewhat different situation. The table below gives results from an impact tester on paper wood sandwiches. The results are expressed as the difference between sapwood and heartwood on each sample. There are 24 such differences.

Paper wood Sandwich
Impact Tester

Sapwood-Heartwood	Rank	Sapwood-Heartwood	Rank
0.07	9.0	0.11	15.0
0.13	17.5	0.09	11.0
0.07	9.0	-0.02	- 2.0
-0.01	- 1.0	0.71	24.0
-0.10	-13.0	0.30	23.0
-0.14	-19.0	0.25	22.0
0.05	6.0	0.07	9.0
0.03	3.0	0.06	7.0
-0.12	-16.0	0.04	4.5
0.04	4.5	0.10	13.0
0.17	20.0	0.13	17.5
0.10	13.0	0.23	21.0
			+249
Probability = 0.01 for - 61			- 51

These differences have been assigned rank numbers disregarding the signs of the differences, and then the ranks have been given the same sign as the differences from which they are derived. The sum of the positive ranks is 249, while the sum of the negative ranks is -51. If the mean difference between sapwood and heartwood were zero, we would expect the positive and negative rank totals to be about the same.

It is possible to calculate the probability of obtaining by chance

a total of one sign as low as 51, by enumerating the number of ways of making up all totals from 0 to 51. This number must be divided by 2^{24} which is the number of ways of getting all possible totals. The result multiplied by 2 gives the probability of getting a + or - total of 51 or less. Tables are available (9) which give the critical totals corresponding to probabilities of 0.05, 0.02, and 0.01. According to these tables a total of 61 would correspond to a probability of 0.01, and therefore the observed total of -51 must be less probable than .01.

With as many as 24 observed differences it is possible to approximate quite closely the true probability by assuming that the distribution of rank totals is normal, with a standard deviation of $\sqrt{(2N+1)\overline{T}/6}$, where $\overline{T}$ is the expected rank-total of one sign, under the hypothesis that the mean difference is zero, while N is the number of paired differences to be ranked, in this case 24. The expected total $\overline{T}$ is one-half the sum of the numbers 1 to 24 or 150. The standard deviation is found to be 35. The deviation from expectation is 99. The ratio 99/35 or 2.83 corresponds to a probability of 0.0046.

The rank Tests described above may be generalized to deal with the comparison of more than 2 categories or groups. As an example we may consider the following table which shows carbon yields in a catalytic cracking pilot unit for different numbers of cycles per test period(2).

Carbon Yields

Cycles/Test Period	8	rank	12	rank	24	rank	32	rank
Period 1	4.28	(2)	4.34	(3)	4.84	(4)	4.15	(1)
" 2	4.37	(3)	4.35	(2)	5.18	(4)	4.21	(1)
" 3	4.25	(1)	4.35	(2)	4.43	(4)	4.39	(3)
" 4	4.40	(2)	4.11	(1)	5.15	(4)	4.59	(3)
" 5	4.54	(3)	4.38	(2)	4.85	(4)	4.35	(1)
" 6	5.19	(3)	4.38	(1)	5.24	(4)	4.60	(2)
	27.03	(14)	25.91	(11)	29.69	(24)	26.29	(11)

If the values are ranked for each period with ranks 1 to 4, the ranks for each column may be totalled. A quantity Chi-squared may be calculated from the rank totals by the following formula:

$$\text{Chi-squared} = \frac{12}{n\ p\ (p+1)} \times \Sigma(T)^2 - 3\ n\ (p+1)$$

Where T is a rank total, n is the number of rows, and p the number of columns.

On substituting the proper numerical values in this expression, Chi-squared is found to be 11.4, with 3 degrees of freedom, one less than the number of columns. The probability of obtaining such a value by chance if the number of cycles per test period were without influence on the carbon yield is only 1 in 100. It is therefore justifiable to conclude that such an influence exists.

Usually the experimenter will not be satisfied to learn merely that the cycles per test period have an influence on the carbon yield. He will wish to make individual comparisons to find out which categories differ from which. This may be done by calculating a difference D between rank totals such that any two differences which exceed this value may be considered to differ from each other with 1 chance in 20 of being wrong.

The formula for D^2 is as follows:

$$D^2 = (n)(p)(p+1)(q)^2/12$$

where q is taken from a table of the Studentized range, and the table is entered in the column headed 4 since we have 4 categories or groups, and the row is indicated by the sign for infinity, and is the bottom row of the table.

In the present case D is found to be 11.4, and it may be concluded that 24 cycles per test period gives a significantly higher carbon yield than 12 or 32 but not significantly higher than 8.

The example given above is a case of a "two way" classification since the measurements are classified in two ways. The columns represent cycles per test period, while the rows represent test periods.

In case we have to deal with a one-way classification, a modified form of the rank method has been described by Kruskal and Wallis (4). In this case all the measurements are ranked from low to high, and chi-squared is calculated from the following formula:

$$\text{Chi-squared} = \frac{12}{N(N+1)} \underset{1}{\overset{c}{\text{Sum}}} \frac{Ti^2}{ni} - 3(N+1)$$

N is the total number of measurements, C is the number of treatments or groups being tested.

The rank totals from each column are squared, divided by the number of items in the column, and the results summed. When this method is applied to the previous example a chi-squared value of 10.54 is obtained which is not very different from the value obtained previously of 11.4. This is because the classification by rows is of little importance in this particular case.

There are certain properties of tests of significance which are important to investigate before adopting some proposed test in place of one which is well established. One of these properties is consistency. A consistent test, roughly speaking, is one which is more and more likely to give the right answer as the sample size is indefinitely increased. Suppose we are comparing two groups by rank methods with a number of replicate measurements for each group. The measurements from one group are labelled x and those from the other y. Suppose we wish to test the hypothesis that there is an equal chance for an x to be greater or less than a y against the alternative that the probability of an x being less than a y is not 1/2. It has been shown that the unpaired rank test previously described is consistent against this alternative (5).

Another important property is the power of a test against a particular alternative. The alternative usually of interest is one in which two populations have the same variance but may differ in their means. It is convenient to consider the <u>power efficiency</u> of a test, which is determined by the relative number of measurements required by the test being considered compared to the number required by the t test to achieve the same power.

It has been shown (1) that the power efficiency of the unpaired rank test relative to the t test lies between 93 and 96 per cent under conditions most likely to be of interest. The asymtotic efficiency as the number of measurements is increased without limit has been shown to be

3 divided by pi, or about 95.5%.

The rank tests described above have been illustrated by experiments of rather simple design, but rank methods may be used in more elaborate designs. For example, in a factorial design with the factors at 2 levels, each contrast consists of a set of paired comparisons, and the paired rank method described above may be used in determining the significance of differences. A 3 factor experiment at 2 levels may be laid out in an 8x8 latin square design, where the rows and columns of the square represent variable conditions which we wish to prevent from influencing the conclusions about the factors of interest. Even in this rather complex situation, rank methods may be used to test significance, and the conclusions are independent of row and column effects.

The following diagram shows a 3 factor 2 level experiment in a latin square:

(1)	a	b	c	ab	ac	bc	abc
a	(1)	ab	ac	b	c	abc	bc
b	ab	(1)	bc	a	abc	c	ac
c	ac	bc	(1)	abc	a	b	ab
ab	b	a	abc	(1)	bc	ac	c
ac	c	abc	a	bc	(1)	ab	b
bc	abc	c	b	ac	ab	(1)	a
abc	bc	ac	ab	c	b	a	(1)

A contrast between the high and low levels of a could be made up of 4 differences of the type a-(1), 4 of the type ab-b, 4 of the type ac-c, and 4 of the type abc-bc, or 16 in all. These differences may be taken in such a way as to be independent of row and column effects. For example the sum of the values for a's in column 2 row 1, and column 1 row 2, compared with the sum of the values for (1) in column 1 row 1 and column 2 row 2 is necessarily free of any row or column effect. The same holds true for the remaining contrasts. The significance of the contrast may be tested by the paired rank method, assigning ranks to the differences. One would of course randomize the rows and columns of this square before the experiment.

Literature Cited

(1) Dixon, W. J.
"Power under Normality of Several Nonparametric Tests"
Ann.Math.Stat. 25, pp. 610-16, 1954

(2) Huhndorff, Roland F.
"Statistical Procedures applied to a Pilot Unit"
Ind.Eng.Chem. 41, pp. 1300-1303, 1949

(3) Kendall, M. G.
"A New Measure of Rank Correlation"
Biometrika 30, p. 81, 1938

(4) Kruskal, William H., and Wallis, W. Allen
"Use of Ranks in One-Criterion Variance Analysis"
J.Am.Stat.Ass. 47, pp. 584-621, 1952

(5) Lehmann, E. L.
"Consistency and Unbiasedness of Certain Nonparametric Tests"
Ann. Math. Stat. 22, pp. 165-179, 1951

(6) "Student"
"An Experimental Determination of the Probable Error of Dr. Spearman's Correlation Coefficient"
Biometrika, 13, pp. 262-282, 1921.

(7) White, Colin
"The Use of Ranks in a Test of Significance for Comparing two Treatments"
Biometrics 8, pp. 33-41, 1952

(8) Wilcoxon, Frank
"Individual Comparisons by Ranking Methods"
Biometrics Bulletin 1, pp. 80-83, 1945

(9) Wilcoxon, Frank
"Some Rapid Approximate Statistical Procedures"
American Cyanamid Co., New York, 1949

ELECTRONIC DATA PROCESSING

J. D. Stevenson
Hughes Aircraft Company

Much has been said and written on the subject of processing Quality Control data through use of a punched-card system. Various authorities have published theoretical approaches to and have prognosticated the advantages of such a program. This is a description of a practical system and its formation which is in operation;and though still plagued with many defects, both technical and operational, it is a vast improvement over previous "long-hand" methods. As a result of this program, new horizons have been opened which were previously inaccessible due to the exorbitant cost of manual data reduction.

In establishing such a system, the first problem is one of economics. A system must indeed show a substantial increase in Quality Control efficiency in order to justify in the minds of Plant Management the initial expenditure required and an increase in overhead costs. A practical survey approach is as follows:

1. Itemize each step in the complete existing system of data gathering, reduction, analysis and reporting operations. Estimate the operating costs of each.

2. Determine those functions which may best be performed by machine, and estimate the costs of those functions.

3. Dovetail the proposed changes into the over-all system and test each phase for inconsistencies.

4. Eliminate all "nice but not necessary" features and take advantage of any cost-reducing innovation.

5. Estimate the value of any features resulting from the newly acquired data handling mobility.

6. Draw up a balance sheet showing relative costs and values of the two programs.

With this mass of factual data, the next problem is one of salesmanship.

The punched card system should perform the operations of compilation, collation, reduction and reporting, since it is in these areas that machines excel. This leaves program planning, report analysis and corrective action feedback as the prime functions of the Quality Control Engineer.

The most critical element in a punched-card system is the transcription of measurement data into punched card form. It is here that the greatest error in the entire program will occur. If the method of entering data in machine-usable form is incomplete, restrictions will exist throughout the remainder of the system; if in error, inaccuracies will result. It is here, also, that human error is present.

The punched card system should enter the program as close to the actual measurement point as possible in order to avoid any errors in

unnecessary transcriptions and human calculations. Of the many possible ways to enter data on a punched card, there are three worthy of consideration:

1. The existing inspection record form is maintained and data is transferred by hand to a form from which keypunch may be performed.

2. The inspection record form is changed to a form from which keypunch may be performed directly.

3. A mark sense card is used as the inspection record form and a mark sense machine is used to punch the card.

The first method compounds human error and should be avoided. The other two methods are in use and appear satisfactory.

Much planning time is necessary on formulating means of coding data so that machine functions may be accomplished economically and at the same time using codes which are relatively simple for encoding and analysis.

At Hughes Aircraft the problem was a particularly difficult one. Upwards of 200 unique electronic assemblies are simultaneously produced with an average of about thirty variables to be measured and 200 attributes to be inspected per unit. Since, in general, all units are inspected and tested and no known defect is allowable, then the inspection records consist of:

1. A record of the rework necessary to bring an assembly within specification limits; and

2. A record of the within-specification measurements of all variables and a certification that all attributes have been inspected and are within specification.

The first record has been arranged so that keypunch may be performed directly. All codes used to describe the defects are established in such a fashion that the inspector lists his defects encoded and in the order desired on the keypunch card. The keypunch operation is nothing more than copying directly. The following items are listed on the punched card with one card per defect used:

1. Assembly identification.

2. Nature of the defect.

3. Severity of the defect (how does it affect assembly function?).

4. Cause of the defect (department responsible).

5. Physical location of the defect in the assembly.

6. Location of the inspector who discovered the defect (at what point in the manufacturing process?).

7. Vendor name (if purchased part).

Tabulations are made as follows:

1. A component summary which lists defects by component application in such a fashion that problem areas are highlighted for engineering investigation and corrective action.

2. A manufacturing defect summary which highlights by way of a demerit rating system those problem areas in the factory which build defects into the product.

The variables data are recorded by pencil directly on a mark sense card. The IBM mark sense card has a recording area of twenty-seven digits or a total of two hundred and seventy unique recording spaces. Due to limitations in normal machine installations, the equipment is incapable of handling in excess of two entries per line or a total of fifty-four entries per card. Since the average unit required some thirty entries in terms of so many volts, ohms, etc., and it was explicitly desired to normally have not more than one card per assembly, the problem was a knotty one. The solution here again was in selecting a suitable code.

The inspection procedure for each assembly in its original form called for a given measurement and listed the specification limits for each variable to be investigated. These procedures were revamped so that the area between specification limits was divided into five equal parts numbering 1 through 5. As an example, if the original read:

Step 16. Measure voltage at test point seven. Should be 100V. ± 10% reading ________.

It was changed to read:

Step 16. Measure voltage at test point seven. Mark the space on the IBM card corresponding to the numbered group in which the measurement falls:

1. 90.0V to 93.9 volts
2. 94.0V to 97.9 volts
3. 98.0V to 102.0 volts
4. 102.1V to 106.0 volts
5. 106.1V to 110.0 volts

This plan was carried through for all steps of all assembly inspection procedures so that the results of measurement of any variable could be recorded as a number in the series 1 through 5. Next, the card to be used as the record form was divided so that it has fifty-four data recording areas, each one of which contains five spaces numbered 1 through 5. These areas were then numbered 1 through 54. Thus all variable measurements of any assembly could be entered on one card provided the total number of such variables did not exceed fifty-four.

A peculiarity of the IBM card is that although it provides space for only twenty-seven digits of mark sense information, up to eighty digits of information may be punched into the card. Thus if a card is

This picture illustrates the use of a mark sense card in recording inspection variables. The card holder is a devise used as an aid, by the inspector, in placing his data in the proper lines and columns. The loose leaf card cutouts shown have two variables and their cells printed on each card. The cards are so cut on their right edge that as they are turned down over the mark sense card, the right edge aligns with the proper line on the card.

filled out in its entirety on the mark sense columns; then when it is punched this information will be placed on only twenty-seven of its eighty possible digit lines leaving fifty-three other lines available for keypunch information.

This peculiarity was turned to good advantage by utilizing it for "header" information. Prior to recording data on the card via mark sense pencil, the inspector or in this case electronic tester, if you wish, records in non-conducting ink such information across the top of the card as:

1. Assembly identification number (or code).

2. Assembly serial number.

3. Date.

4. Inspector identification number.

5. A code indicating the point in the manufacturing process where the inspection is being conducted.

6. The number of the specification to which these particular variable measurements are referred and latest engineering changes thereto.

7. Whether inspection is an original or the result of a previous rejection.

After completion by the inspector, the cards are routed immediately to a control point where they are visually checked for completeness and condition. If any card is incomplete, smeared or otherwise mutilated, it is returned to the tester for correction. Accountability is maintained at this point to guard against lost or misplaced cards.

From the control point, the cards are sent to the machine processing center where they are mark sensed by machine. This mark sensing consists of reading the data appearing on the card in pencil and permanently punching this information into that same card. This is done as soon as possible since time and handling tend to smear the cards, resulting in mispunches and therefore erroneous data. Following mark sense, the cards are given to keypunch operators to punch the "header" information into the body of the card. The cards are verified and then stored until a report is desired.

Once each month the cards are collected, collated by manufacturing assembly number and inspection step number, and summarized by cell divisions. These summaries are then tabulated in a report.

The example on the following page will show the form of the report: If 100 units of a given assembly were inspected during the month and the inspection step 16 summary showed that 80 units were measured in Group 1, 10 in Group 2, 5 in Group 3, 3 in Group 4 and 2 in Group 5, the report would show:

UNIT 386				JAN. 1955	
TEST NUMBER	CELL				
	1	2	3	4	5
1	18	20	24	20	18
15					
16	80	10	5	3	2
17	10	23	34	23	10

This information appears for each inspection step of each inspection procedure in use. Briefly this represents a histogram for each variable for the production of each unit during that period, if it is kept in mind that the data represented is only "good" data and does not include those measurements which were made that resulted in rework to bring the assembly to a within-specification status.

A theoretically sound analysis of a histogram drawn from such data is almost an impossibility, but a thoroughly useful analysis for corrective action is not difficult. In other words, the histogram cannot be used as a measure of quality but is extremely useful in highlighting troublesome areas from an engineering, inspection and manufacturing viewpoint.

In the example given, the distribution is obviously skewed, but the next question in analysis is "what constitutes a good or bad distribution?". Since most electronic measurements concern the results of the complex interaction of many component parts, each of which varies within its own specification limits and almost all of which were manufactured and purchased through use of sampling plans and separately established AQL's, the build-up and cancellation of individual tolerances may result in an infinite variation in measurement readings. Unfortunately it is almost impossible to observe or calculate all of these possible effects in the design of an electronic circuit; consequently, many unusual and quite often surprising effects result. After many consultations with design engineers and quality control mathematicians surveying reject data over a long period of time and much "coin-flipping", it was decided that although such is not the average case by quite some margin, it should not be unusual to have a rejection rate of 10% on any one particular function. At least this would be a good point to start from and could be modified according to the work load on the Corrective Action Unit.

Assuming a normal distribution, a symmetrical 10% rejection and a five cell division, the marginal cells (1 and 5) should include somewhat less than 12% of the units each. The present analysis makes use of this figure. Each month when the tabulation is received, an analyst checks each step of each assembly. Any step that exceeds 12% in the columns 1 or 5 is noted and the information is forwarded to the Corrective Action group for investigation.

The Corrective Action group compares the inspection data analysis step by step with the rework tabulation. Normally a marginal condition in the analysis is verified by a high rework figure in that same functional area. When such is the case, the method of measurement is verified and the following courses of action are taken:

1. If the curve is symmetrical but marginal, the component or components experiencing high rejection rates and immediately adjacent components are investigated to prove they are within their individual specifications. If such is found to be the case, it may be found that the original functional specification may be unnessarily rigid. If upon investigation it is found to be a reasonable and definite requirement, the specifications of the components used to develope the function are investigated to determine if they are sufficiently close and complete. If no error is located at that point, the problem is referred to design engineering for a basic change.

2. If the curve is marginal and skewed, the same procedure is used except that it may be found the basic specification is unreasonable and can be shifted over the experienced distribution.

The present basic program will experience many changes in the future. It is planned that it will be expanded to cover all phases of inspection and undergo considerable analysis with consequent greater utilization. In its present form this program has all of the advantages and possibilities which were expected of it. Its installation, though plagued with innumerable minor problems proceeded almost according to original schedule and has proved to be a highly interesting and profitable venture from a Quality Control efficiency standpoint.

"COMPETITIVE QUALITY - MOUNTAIN OR MOLEHILL?"

R. H. Lace
Riverside Paper Corporation

Although there may be special reasons at times for evaluating competitive quality the following are the four basic purposes for which evaluation is made:

1. To inform the Sales Department and the men in the field, so that they know what to do and say when the question of quality in competitive products is brought up. Specific facts are always better than vague generalities. Weaknesses of competition can be emphasized and strong points "soft-pedalled."

2. To inform the Development Engineering and Research Departments about the quality of current competitive products so that planning for product improvement can be adjusted to conditions in the field as they arise.

3. To inform the Production Departments of the relationship of their outgoing product quality to that of competition, not only on the basis of "Squawks" from the field, but rather on the basis of a factual evaluation which indicates points for improvement within the framework of present specifications.

4. To aid in Management decisions which concern the relationship of a particular company to others in the field, or to specific products in a given sales program. Facts as to quality are necessary as well as facts in regard to financial strength, sales effectiveness and organization.

It will readily be recognized that the preceding purposes require the same type of procedure for determining quality that is used in the quality audit which many firms use as a part of their quality control program. Here extremely small samples of outgoing production are evaluated for quality in total and all other available data in regard to field performance is accumulated and analyzed.

Many sources are available to us insofar as data for the evaluation of competitive quality is concerned. Our salesmen are usually very voluble as to the effect of competitive quality on their sales volume. Where the product is such that either a field service organization is maintained, or where the distributor maintains competing products, additional information on competitive quality is often available. Analysis by the quality control organization of competitive products is the final, and can be the most reliable, source of information.

We say "can be" since the possibilities for biased opinion are always present, and must be reduced to a minimum through the proper procedures, in order for the final report to be most effective.

One illustration of the difficulties of relying upon field data as such, can be gotten from the following chart which appeared in Business Week Magazine recently as part of its consumer motivation series.

(See Chart #1)

In this case an evaluation of competitive product quality might have disclosed the information that it was necessary to obtain through an extensive consumer preference analysis. While the review and analysis were made with the end of determining whether or not the advertising program of the company was effective, the facts which resulted in a shift in advertising emphasis would have been at least partially (if not wholly) unveiled through a competitive product quality evaluation.

As will be noticed, the characteristic of the product which customers felt was most important was strength, and those who deserted the competitive product did so primarily because of strength limitations. It is, of course, true that while the basic quality factors could have been determined, the reasons for purchasing might not have been as easily discerned.

If getting the facts about competitive quality is so important, what procedures must we follow to insure that we are actually getting the maximum amount of practical factual information?

First, we must know the effectiveness of a single purchase sample. Its ability to give us specific information about the quality of the competitive product as it exists in the sample is important, but it can, also, give us current information about the production process from which the sample came. In other words, we can know that the sample is indicative of its own quality, and can also know how the quality of material being currently sold is related to it, and can predict how the quality of material currently being manufactured is related to it.

In the case of one of our customers, the Lincoln Paper Company, and its parent company, DITTO, Incorporated, when a new product is put into the market by a competitor, or when they are evaluating the competitive market quality status in relation to a new product they are about to introduce, or if they are determining the continuing quality of competition in relation to their own current product, they use knowledge of production processes in making decisions as to how much and when to buy competitive products in order to come up with an adequate picture of competitive quality.

In the case of a product like an office machine, we know that basic design usually changes slowly while minor modifications are frequently in process. We must, therefore, consider this in making our evaluation where we find that there are points of obvious inferiority in a product. This is true to a somewhat lesser degree in paper products, where improvement must always be considered in relation to existing inventories and where changes are made slowly except in the case of extreme quality failures. In the case of most supply products where use habits are of concern, the development of uniform appearance, feel, texture, and odor is quite important to the manufacturer and a small sample can give a good

indication of the overall product quality.

Where, because of the nature of the materials and processes used in the industry, wide variation in product quality exists, the isolated purchase of a product will give information, but it is correspondingly more difficult to relate average competitive product quality to that of the sample.

Thus, in the field of office systems and general duplicating equipment, competitive quality evaluation is based on samples ranging from one each of the basic machines in an important competitive line to a single machine of some other competitor, independent of the number of machines which may be in the line.

On paper and operating supplies, purchases may be anywhere from quarterly to annually, depending upon the importance of the particular competitor and/or the possibilities of changes in the product as a result of efforts toward improvement.

How do we analyze the quality of competitive products? Exactly the same way as we would analyze the quality of our own product. We first make an overall determination of the things that the product is sold to do. Here obvious inadequacies can frequently result in reducing the amount of effort necessary for the evaluation. The product may be so inferior that it does not present a potential competitive threat, because it is not designed to do the job which "field needs" analysis shows is necessary for the particular quality and price range.

Also, a situation could arise where the product will not do the job it is sold for as the result of defective material having inadvertently been shipped, rather than of design or specification qualtity being decidedly below field needs. This will also be determined during the analysis of the product.

Now we come to the area that is most important: Determining the relationship of our design or specification quality, to: competitive product quality and to: actual consumer needs. These can be shown graphically as follows:

(See Chart #2)

A. Here we have a situation where competitive quality is below our design quality while both are above the minimum field needs. This represents a situation where our efforts, as a result of the evaluation, should be to stress the quality features and possible aesthetic values of the products in our sales and advertising approaches without further effort being required by the production or development group.

B. Here is a situation where our product is no better than that of competition and both have the same relationship as to needs of the field. Here it would seem that further effort by the engineering or development group is necessary in order to improve the quality of the product, while other characteristics, such as service organization, should be

stressed for the present in either advertising or sales.

C. This is a situation that the sales division sometimes yells about, but that a good salesman likes to sink his teeth into, where competitive quality and our own design quality are above field requirements but where competitive quality is above ours in total. Here we have a situation where the design group needs to frantically get improvements underway while the sales department holds the line with either stress on service organization or other semi-intangibles.

D. Here we have a situation that could really be a problem, where the quality of the competitive product is above our design quality, and where the minimum field needs are in between: i.e. our own product falls short of meeting the requirements of the customer. Here our best move is normally to back up the efforts of our sales division through a company-wide product improvement drive.

I say "Normally" since later on I will give an illustration of the type of problem that can arise where the quality control department cannot make the complete recommendation, since factors other than quality must be considered in a product improvement decision.

There are two methods in current use for arriving at the total quality rating of designs (or specifications) and products. These are:

1. A merit rating system.

2. A demerit rating system.

In the merit system the product specification requirements are analyzed and values assigned to the individual characteristics on a weighted basis so as to emphasize those characteristics which are most important. When the product is evaluated, each characteristic receives a percentage of the total possible points; these are then summed up and the ratio of the points achieved to the total possible points is the quality rating of the product.

As an example of the use of this type of rating, the procedure used for hectograph cleansers or cream soaps may be cited.

(See Chart #3)

In using the demerit system the presence of undesirable characteristics is noted with a weighting being assigned to each on basis of its effect on either the customer, the products with which it will be used, or performance of the product itself. The Western Electric Company pioneered the use of a weighted system of this type and it is still in use in their quality audit procedures. Weightings will vary from 100 for presence of a characteristic which will cause complete failure of the product or which will endanger life or health of the user, or which will cause the failure of material with which it is to be used, to from one to ten demerits for imperfections which the average customer will normally not notice.

As an example of this rating I have a carbon and Masterset evaluation form.

(See Chart #4)

There are applications in which both types of ratings appear to be necessary, but the difficulties of a summarized report containing such diverse factors are quite obvious.

The problems inherent in using the systems described above are several:

1. Either an individual or a committee must determine the weight to be assigned to each product characteristic or deficiency.

2. The figure or graph by itself does not indicate the corrective action necessary nor does it indicate specific points of superiority, and must be supplemented by such information.

3. In the case of complex products such as office machine, the problems of merit rating also become quite complex; it is better to use the demerit system in combination with a great deal of objective (or nearly so) reasoning on what the customer really needs and the degree of importance of specific characteristics.

These difficulties are normally not such as to cause an impossible situation, but rather must be considered in planning prior to application of the procedure.

If the planning is properly conducted and the weightings fairly arrived at on the basis of a representative group decision, (sales, development, production, quality control) then this type of analysis can be of considerable help in relating your gradual product improvement to that of competitive efforts. Customers' needs, as these are affected over a period of time through either increased knowledge of them or because of adversiting stress, can also be evaluated and compared.

Earlier we indicated that, while the report on competitive quality is the responsibility of the quality control department, and that specific recommendations for corrective action may be issued, the basic responsibility for action is not that of the quality control department, but rather top management, since many factors other than quality as such must be considered.

As an example, some time ago they analyzed a competitive liquid soap, and found that it was not only superior to their own product, but, as a result of its introduction, the relationship to field needs was modified so that the product no longer was adequate. This normally would have called for a review and a decision for product improvement, with either a specific assignment given to the research and development group, or an arrangement made for manufacture and/or distribution of the better product.

However, a complete analysis of the situation indicated that while the product was superior, the company distributing it was not only cold to anyone's distributing their product on a royalty basis, but they were determined to go it alone without adequate finances or distributor

organization.

A decision was made to do nothing about the immediate situation, while the old product was put on a long term improvement basis (this being an accomplished fact, it is safe to report the foregoing). As anticipated, there was no effect on sales volume, and the current product quality position has been strengthened.

We may summarize the foregoing as follows:

1. We have a great many sources of information about competitive product quality as well as our own product quality.

2. The bits and pieces of information need to be tied in with an objective product quality evaluation, so that a complete relationship may be established.

3. The divisions affected, as well as top management, need to know the results of the objective evaluation, together with specific recommendations for such action as may be necessary to maintain or improve field position insofar as product quality is concerned.

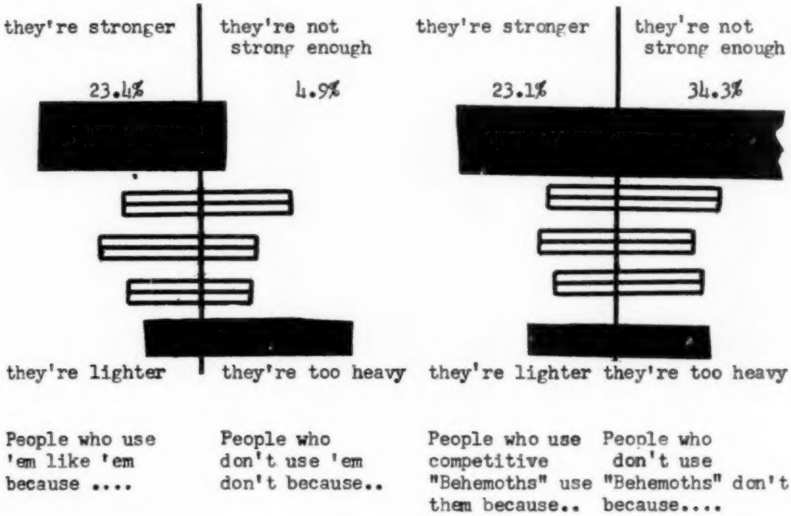
4. Either a positive approach through a weighted merit rating may be used, or the somewhat negative, but more easily administered, procedure of demerit rating and weighted evaluation can be used.

5. In either case the weightings and procedures must be firmly established beforehand by a representative group including the sales, engineering or development, manufacturing, quality control and sometimes financial or purchasing divisions.

6. The final report can be a means towards maintaining and improving product quality with added sales volume and better net profits.

WHAT FEATURES REALLY SELL
YOUR PRODUCT?

Here's how Ajax "Widgets" found out through a
consumer motivation survey -



c
BUSINESS WEEK

Chart #1

PRODUCT AND CONSUMER QUALITY
RELATIONSHIPS

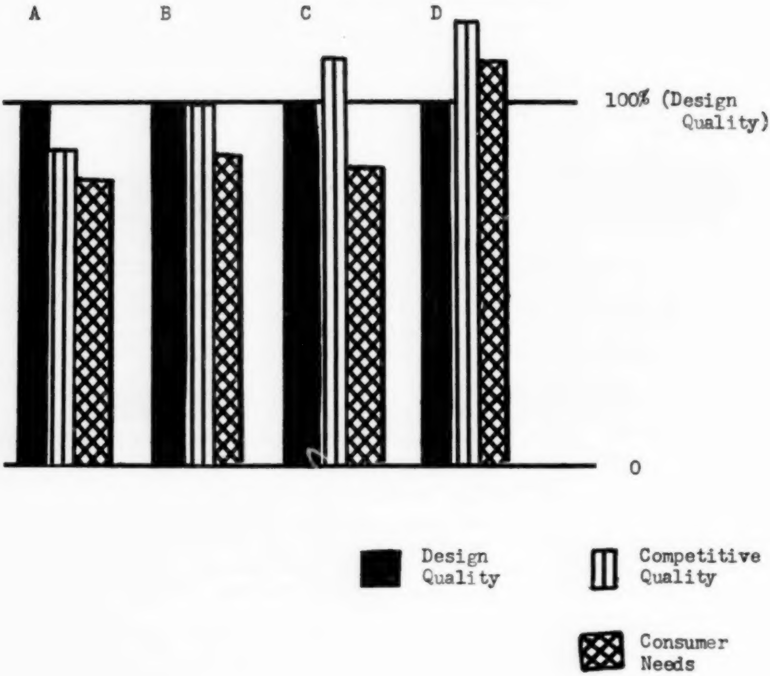


Chart #2

CREAM SOAP SPECIFICATIONS

I. GENERAL CHARACTERISTICS

A. Color - 25
B. Perfumed- 25
C. Texture

1. Smooth. Cone penetrometer reading at F. shall be 25
2. No solid particles shall be felt in any sample rubbed between the hands. 50

D. Heat Resistance

50 1. Shall not pour under F. (Any 3 or 5 samples per lot)
100 2. Shall not precipitate or separate upon cooling and re-solidifying after melting. (Any single sample) Note:- Pour point: temperature at which the soap will immediately flow out of the can when it is gently laid on its side.

II. PERFORMANCE

100 A. No more than 10 grams of soap shall be required to clean hands prepared in standard manner. This quantity may be in two portions.

50 B. Soap removal from hands should be complete within 2 min. when removing from hands prepared as in 2-a, immersion under tap of ½" opening, about 5 gpm, water at 110° F.

50 C. Six washings per day for three working days when ambient Relative humidity is 30% minimum and shall not cause chapping or drying.

25 D. No stickness or slipperiness shall remain after removal of soap as in 2-b, followed by drying with absorbent cloth.

III. PACKAGING

A. 6 oz. Tubes (Collapsible)
B. 1 lb. Glass Jars or Metal Cans
C. 6 lb. Glass Jars or Metal Cans
D. 25 lb. Metal Pail

IV. LABELING

A. Labels and containers shall be approved by Packaging Committee.

PHYSICAL QUALITY STANDARDS FOR CARBON

CHARACTERISTICS	DEMERITS	CHARACTERISTICS	DEMERITS	CHARACTERISTICS	DEMERITS	TOTALS
		-- Defects due to formulating and coating --				
Uncoated areas	100	Scratches	50	Varicolored	10	
flaking	100	Ridges	50	Coated dimension 1/32 out of tolerance	10	
lines in carbon	100	Lumps	50			
lines on edges	100	Foreign Material	50			
Coated dimensions 3/32 out of tolerance	100	Dirty Cleanedge	50			
Wrinkled tissue	100	Coated dimensions 2/32 out of tolerance	50			______
		-- Defects due to processing coated carbon --				
Collated tissue	100	Creased tissue	50	Cutting dimension 1/32	10	
Wrinkled carbon	100	Cutting dimension 2/32 out of tolerance	50			
Dirty master	100					
Torn master	100					
Creased master	100					
Incomplete spraying of edge	100					
Cutting dimension 3/32 out of tolerance	100					______
		-- Defects due to handling and storage --				
Broken box	100	Soiled box	50	Bleeding or mottling of 10-15%	10	
Bleeding or mottling of 30-100%	100	Box with corner scuffed	50			
		bleeding or mottling of 15-30%	50			

TOTAL DEMERITS ______
SAMPLE SIZE ______
RATING ______

______ COMMODITY NUMBER ______ COMMODITY ______ DATE ______ WHERE CHECKED

CHECKED BY ________________

VISUAL INSPECTION OF CLOTH

George W. Haynes
Avondale Mills

In 1950 our President, Mr. J. Craig Smith, decided to assure our customers that they could always count on Avondale Mills' cloth as being "Top Quality." He gave this assignment to Mr. Gardner Hailes, Quality Control Manager, to work on. After Mr. Hailes got into the details of the problem he asked himself this question - does our inspection department actually control our outgoing quality of cloth or does the weave room? This is a very good question to ask whenever the plant manager is in charge of inspection as well as manufacturing which is the case in Avondale Mills. With this thought in mind, the point grading system and the control for it was designed.

After working out the point grading system we ran into the problem of selling it to our foreman of the inspection department. Our foreman had been with us over ten years and was used to scanning cloth to determine whether it was first or second quality. He not only felt strongly about his ability but the ability of his inspectors. It was very diffi-for him to realize that his inspectors could not grade as well as he since he had been over every type defect with them so many times, but this was not our real problem. For some reason, the foreman felt that the Quality Control Department was going to take over the inspection department. Even though the Q. C. Engineer told him many times he would not believe it. Finally, this was discussed with the Plant Superintendent and he in turn reassured the Inspection Foreman that he did not have anything to worry about. After the Foreman thoroughly understood that his job was safe, he was most receptive to the point grading system. Then the training of individual inspectors was started.

For simplicity sake we will take one plant and follow the inspection procedure through. For our example, we will use the Birmingham Plant where we make 250 to 500 yards per pound goods, 36 to 54 inches in width. All goods are finished in another plant. All fabrics are inspected and burled in the grey at the plant and graded on the folder in the finishing plant for finishing damages only.

As you read the mechanics of our point grading system, please notice that it was designed to protect our customers and yet provide a systematic method for our inspection department.

POINT SYSTEM OF GRADING

I. General Considerations:

1. There are two main considerations in grading cloth:
   A. The frequency of defects present.
   B. The seriousness of the defect.

2. The seriousness of the defect is determined by two principle factors:
   A. The intensity of the defect.
   B. The size, or length of the defect.

3. Intensity (or "obviousness" or "visibility") affects

whether a cutter will see the defect or not, and if he does, whether he will cut it out or cause a defective ("second") garment, depending on his practices.

4. Length also affects the obviousness of a defect, and in addition, it determines how many panels or pieces (and hence garments) may contain the defect. Therefore, length is more important than intensity.

5. The simplest possible system that will take both of these factors into account, will provide for a two-notch breakdown for each factor.

6. Let us divide intensity into two categories:
   A. "Minor" (crudely defined as "obvious").
   B. "Major" (crudely defined as "very obfious).

7. Let us divide length into two categories:
   A. "Short" (up to 6" long).
   B. "Long" (6" to 18" long).

8. In order to use the smallest possible numbers, let us charge one point for the least serious defect. (A "short" "minor" defect - this is the basic unit defect).

9. When we increase the intensity only of the basic unit defect, let us double the number of points, and charge two points for a short "short major."

10. When we increase only the length of the basic unit defect, let us triple the number of points and charge three points for a "long minor" because length is more important than intensity.

11. When we increase both the intensity and length of our basic unit defect, let us quadruple the number of points. In other words, charge 4 for a "long major."

12. When the above is tabulated we have the following:

AVONDALE POINT SYSTEM OF GRADING

	Minor	Major
Short (0 to 6")	1	2
Long (6" to 18")	3	4

II. Definitions of Defects:

1. Sub Minor: A defect which is not obvious, may not be notic- at first glance, and would not be likely to cause a garment so defective that it would have to be sold as a "second" at a lower price. No points are to be charged for these defects, but if a great many of them are present, they should be called to attention, and consideration given to grading the entire cut of cloth as seconds.

2. Minor: A fairly obvious defect which is noticeable more or less at first glance. Might easily cause a defective gar-

ment, charge 2 or 4 points depending on length.

4. Cutting:

A. A hole, split, or broken picks, which might result in the cloth tearing on the tenter frame at the finishing plant, thus ruining 60 yards of cloth.

B. A defect so severe that it would be likely to cause a garment which would not be salable even as a "second." Since such defects are to be cut out, no points will be charged. Flag them with a red string-flag.

III. Advantages of System:

1. It uses the smallest possible points.

2. These numbers are chosen so as to reflect the seriousness of a defect from a cutters view point.

3. It takes into account and provides for the exercise of judgement which will be exercised in any event, whether we recognize it or not.

4. The table checks with common sense.

5. It provides a logical basis for setting up standard samples for reference use by the graders. Such samples will be more consistent and less confusing to the graders because a rational criterion exists for selecting the samples.

6. It provides a means for controlling the "strictness" of the grading without confusing the graders by changing their standards. It is only necessary to change the allowable number of points.

IV. Installing System:

In determining the allowable number of points for first quality cloth. Forty thousand yards of first quality as graded on the old system was taken from inventory and regarded on the point system. A frequency distribution was made from this data. The distribution was schewed so much that it was not practical to compute standard diviation and set the upper limit on average plus three S.D. Instead, the data was ploted on graph paper and the area under the curve was found. The upper limit was set by taking 10% of the area as being out of control. In other words, 90% of the present first quality cloth was assumed acceptable to cutters. We realized in doing this that it was a "dirty method." To play safe, we procured some competitors cloth and inspected it to see if it coincided with our results. Also we checked with several cutters to get their opinion as to how many defects they would be willing to accept and still classify it as first quality.

Standard defects samples were taken from actual production. A

meeting was held with the General Manager, Q.C. Manager, Plant Superintendent, Q.C. Engineer and Inspection Foreman. The minimum intensity of defects were chosen and a point value assigned to them. In explaining this to the inspectors, they were told that if the defect has this intensity or up to the next grouping charge the number of points shown listed. A standard defect sample board was made using minor, major and cutting defects. This was placed near the inspection tables for easy reference.

The Q.C. Engineer trained each inspector for two days. The following was covered. (1) Theory of point grading. (2) Cutters pattern layout. (3) Three thousand yards of cloth was graded. (4) Names of different types of defects. (5) Burl then charge points. (6) She would be the only person inspecting the cloth and would be held responsible for it.

V. Controls of System:

Let us assume that we have the point system in operation. Now our problem is to set up a system to control the inspectors so that they will grade consistently regardless of what quality the production departments produces. As you probably know, when the quality of clothstarts getting bad, an inspector will have a natural tendency to "loosen up" Why should inspectors "loosen up" when the quality goes bad? There are several reasons such as (1) Fatigue - when there are above average defects present in cloth, the inspector has to work harder and it also increases the chances of her missing defects. (2) The Superintendent or Foreman does not want to make a bad record so they put pressure on the Inspection Foreman to "ease up" He in turn will put pressure on his inspectors to "loosen up." Two procedures were set to minimize this sort of thing in visual inspection, they are Evaluation of efficiency of Inspectors and Quality Audit.

VI. Evaluation of Efficiency of Inspectors:

Avondale's evaluation is a composite rating of quality and production.

A random check is made of each inspector by a check inspector. In order to weight cloth with a few number of defects the same as cloth with a large number of defects the deviation of inspector from check inspector in terms of points per 100 yards is used. Then this deviation in points per 100 yards is converted to an arbitrary scale of "0" to 100%.

1. Method of choosing samples:

A small bingo cage with wooden balls in it is used. On each ball there is a number corresponding to an inspectors code number. When a sample is to be taken, the handle on the cage is turned several times to thoroughly mix the balls so as to give a random sample. Whatever number comes up, that is the inspector whose cloth will be checked. The ball is returned to the cage before the next sample is taken. The sam-

ple is taken to the ckeck inspector who regrades the cloth. The check inspector does not see the ticket from the inspector. After the ckeck inspector finishes inspecting the cloth, check inspectors and inspector's tickets are sent into Quality Control. By using a bingo cage it accomplishes two goals: (a) The inspectors never know when they will be checked and (b) The inspectors know that the ckeck inspector is impartial in her selection of samples.

2. <u>Computation of Quality Rating:</u>

Example No. 1:

Cut of cloth is 200 yards long. Inspector A gives 20 points. Check Inspector gives 30 points. Therefore, the difference is 10 points. Ten points divided by 200 yards gives a deviation of 5 points per 100 yards. Referring to Quality Rating (arbitrary) scale, we find Inspector A's quality rating to be <u>84%</u>.

Example No. 2:

Cut of cloth is 200 yards long. Inspector B gives 30 points. Check inspector gives 20 points. Therefore, the difference is 10 points. Ten points divided by 200 yards gives a deviation of 5 points per 100 yards. Referring to Quality Rating scale we find Inspector B's quality rating to be <u>84%</u>.

As the ratings indicate, it is just as undesirable to give too many points as too few.

3. <u>Computation of Precent Production:</u>

Assume standard production to be 6,000 yards for eight hours.

Inspector A's production = 6,000 yards. $\frac{6,000}{6,000}$ = 100%

Inspector B's production = 4,000 yards. $\frac{4,000}{6,000}$ = 66.7%

4. <u>Computation of Efficiency Rating:</u>

Efficiency Rating = Quality Rating Times Production Rating.

Inspector A: 84% x 100.0% = 84% Efficiency Rating
Inspector B: 84% x 66.7% = 56% Efficiency Rating

5. Inspector's A and B were doing an average job of inspection. A was running a satisfactory number of yards for eight hours but B slowed down so much that it was not economical to let her continue to run at this rate. This would be called to Inspector B's attention to help her get straightened out. As a last resort B would be disciplined. The inspector with the highest efficiency rating for the week received a $5.00 award. Her name is put on a board in her department and published in

our company paper. Warning - The Quality Rating System is not effective when there is a large difference in the average points per 100 yards of different styles of cloth.

VII. Quality Audit:

In order to understand the function of the Quality Auditor, it is necessary to know his place in the organization. He reports directly to the Quality Control Manager and the Quality Control Manager reports directly to the President. The Auditor visits each plant once a week. He inspects representative cuts of cloth, going strictly by the established Standards, and compares the results of such checks with the findings of the Mill's Check Inspector. A report is issued to the President, General Manager of Production and the Mill Superintendent. The Foreman is not given a copy because he usually grades the cloth with the Auditor.

If the Mill is "out of control" the superintendent must write the general manager a letter explaining why his inspection department is out of control. The audit is one of the heavy weighted characteristic in the monthly Quality Flag Award. This is an award presented to the mill with the best quality record for the month.

(1) Selection of Auditors Sample:

The Assistant Foreman selects two cuts of cloth each day from the Check Inspector. He does this by placing eight balls, with corresponding numbers to a clock, in a box. For an example, suppose he picks a ball with the number nine on it. He will then get a sample anytime between nine and ten o'clock. He usually operates on the quarter hour so that he can fix a definite time to make his selection. He stores the cloth in the Foreman's office and takes the Inspector's and Check Inspector's tickets into Quality Control Office where they are kept until the Auditor arrives.

(2) The Audit:

Before the Auditor starts his inspections he studies the Standard Sample board, the same one the Inspectors use, to keep himself up to date. He is accompanied by the local Quality Control Engineer and the Inspection Foreman during his inspection. Each cut of cloth is inspected and the results recorded. The rating for the Check Inspector is found the same way that the rating is found for the inspectors. That is, the deviation in points per 100 yards for each cut is found. The average deviation is then computed and then converted to Quality Rating. The reason we do not total Check Inspectors and Auditors points, then get the average deviation in points per 100 yards, is because it would weight the deviation by the number of yards rather than by the number of cuts. We are only interested in how a grader deviates between cuts, because the primary function of an inspector is to sort first from second quality. As you can readily see, the audit will have a tendency to standardize inspection judgement.

VIII. Results:

1. After one year in operation a net savings of $238,000 in labor, off goods and complaints.

2. Reduction in percent seconds - production department has definite standards to meet.

3. Improved moral of production and inspection department.

4. Reduced the number of Inspectors from 23 to 12. This means other savings in dollars and cents in things such as Retirement Fund, Social Security and Insurance.

5. The President of our company has a positive control over outgoing quality.

QUALITY CONTROL AND ITS APPLICATION TO THE BOTTLING OPERATION

W. H. von Meyer
Barry-Wehmiller Machinery Co.
St. Louis, Mo.

Consider the many uses being made today of the statistical quality control programs in various industries. Certainly the time and capital being saved annually by efficient quality control programs warrants the bottler to take a serious look at what is involved in establishing quality control methods in his own plant, and what benefits he may expect once a quality control program has been initiated.

The bottler may be hesitant about becoming involved in a program of which he has not had too much experience or knowledge. He undoubtedly may wonder whether it is the right step to take; he may feel that the benefits gained are dubious; he may lack confidence in the statistical method; the initial investment may be too great; and he may feel that his operation is too small to really benefit by quality control. These are typical of questions which arise in the mind of the average bottler and are serious questions which must be answered positively before any quality control program can succeed in his shop. Such a program, once instituted, will, if the attitude of the management is favorable, succeed.

What, then, is the best way for setting up a control on the overall quality of a product when many factors must be considered in evaluating the final results? The idea of using percent defective will probably be dropped as soon as you have considered the prospect of classifying an entire bottle washing machine as defective! And while variable control charts are wonderfully useful devices, there are many quality requirements that will not fit handily into $\bar{X}$ & R.

When the probability of finding a fault in a product is small in relation to the opportunity for faults to occur, it is possible to use a "C" chart for defects per unit. This comes closer to what we want, because it allows us to group together all of the different kinds of faults into our figure representing each unit inspected. But ordinarily it has the drawback that a minor fault will carry the same weight as a major departure from quality requirements.

The problem, therefore, is to set up a control that will take into account the seriousness of a defect as well as its frequency.

The many independent quality requirements of a product have one characteristic in common; the effect of non-conformance of the user. This point of view allows us to divide the faults found in inspection into several classes graded from least serious to most serious. It is difficult to inaugurate any operation so complicated that its possible faults could not be fitted reasonably into the three following classifications:

1. Very Serious or Critical.
2. Serious or Major A Defects.
3. Moderately Serious or Major B Defects.

When quality requirements have been classified, the relative seriousness of each class is expressed by an assigned weight. The scale of weights used is arbitrary; only the relative weights and the frequency of occurrence for each class of defects will concern us. These weights have been given the term "demerits" and those used in the following discussion are shown on Chart I.

CHART I

EXPLANATION OF DEMERIT CLASSIFICATION

	Mechanical	Electrical	Personnel	Supplies & Services
CLASS I Very Serious Defect. 10 Demerits	Complete failure of operation.	Complete failure of operation.	Major infraction of operating rules.	Complete failure of external supplies or services to line which prevent line operation.
CLASS II Serious Defect. 5 Demerits	Operation erratic or out of adjustment. Requires extensive rework.	Faulty operation or hazardous conditions requiring extensive rework.	Minor infraction of operating rules.	Erratic or defective supplies or services requiring extensive down time.
CLASS III Moderately Serious Defect. 3 Demerits	Operation erratic or out of adjustment. Requires only minor adjustment.	Faulty operation or hazardous conditions requiring minor adjustment.	Moderate infraction of operating rules	Erratic or defective supplies or services causing minor down time.

Let us examine the breakdown of the three classes of defects. From our experience we can reasonably assume that failures in bottling line operation fall into four types which are either mechanical, electrical, personnel, or services and supplies. The seriousness of these failures are subject to change, and it is for this reason we have classified them into three main failure classes; Class I - Very Serious, 10 Demerits; Class II - Serious, 5 Demerits; and Class III - Moderately Serious, 3 Demerits.

It is rather simple to illustrate typical examples of such failures, and for the sake of clarification the following is offered:

Class I - Mechanical - Complete failure of operation. Any operational unit of the bottling must either be functioning or not functioning. A complete shutdown of any bottling unit as a result of mechanical failure would constitute a Class I mechanical defect.

Class I - Electrical - Complete failure of operation. Here again, the complete shutdown of any bottling unit as a result of electrical failure would constitute a Class I electrical defect.

Class I - Personnel - Major infraction of operating rules. A typical example of this defect would be major carelessness of an operator, or complete lack of concern for his duties.

Class I - Supplies & Services - Complete failure of external supplies or services. Should a line be shut down due to a stoppage of steam or water supply or a lack of bottles, cartons, crowns, etc., it should be scored as a Class I defect.

Class II - Mechanical - Operation erratic or out of adjustment. Requires extensive rework. Typical of this may be soaker loaders out of time, labeling out of adjustment, etc. Items in this category cannot be repaired satisfactorily while the unit is operating.

Class II - Electrical - Faulty operation or hazardous conditions requiring extensive rework. An example of this type of defect may be a faulty electrical switch or any electrical function which may present a danger to the operating personnel.

Class II - Personnel - Minor infraction of operating rules. Typical of this defect might be the failure of an operator to make caustic titrations, to send the traveling recorder through the pasteurizer, or a completely unacceptable cleanup job on a unit during regular cleaning period.

Class II - Supplies & Services - Erratic or defective supplies or services requiring extensive down time. Items such as defective crowns, inferior cartons; short failures of steam or water pressures are covered by this classification.

Class III - Mechanical - Operation erratic or out of adjustment. Requires only minor adjustment. Slight mechanical adjustments such as tightening a bolt or adjusting a spring may be classified in this type of defect.

Class III - Electrical - Faulty operation or hazardous conditions requiring minor adjustment. Here an item such as water dripping on a motor, or a temporary electrical adjustment which causes only a minor pause in the production schedule would fall in this classification.

Class III - Personnel - Moderate infraction of operating rules. Typical of this defect might be a minor infraction of operating rules such as general untidiness of working area, or unsatisfactory cleanup of unit during regular cleaning period.

Class III - Supplies & Services - Erratic or defective supplies or services causing minor down time. Items such as an occasional defective crown, or improper glue for the labelers, short duration failure of cartons to the loader or the packer may be classified in defects of this type.

The above is by no means intended to be a complete list of causes for each class of defect, and in actual practice each brewery must decide what its own quality operating level must be.

For the work presented here, it was decided that each bottling line consisted of fifteen separate units. This next slide (Figure 1) copies of which you have received, indicates the units chosen, as follows; (1) Cartons In,(2) Case Unpacker, (3) Soaker Loader, (4) Soaker, (5)Rinser, (6) Soaker Discharge, (7) Filler, (8) Crowner, (9) Pasteurizer, (10) Inspection, (11) Labeler, (12) Packer, (13) Conveyor to Storage, (14) Cartons to Storage, (15) Operators.

The next slide (Figure 2) indicates how the entire fifteen units might appear as a single sample where each operation of the bottling line has four possible sources of defects.

FIG. 1

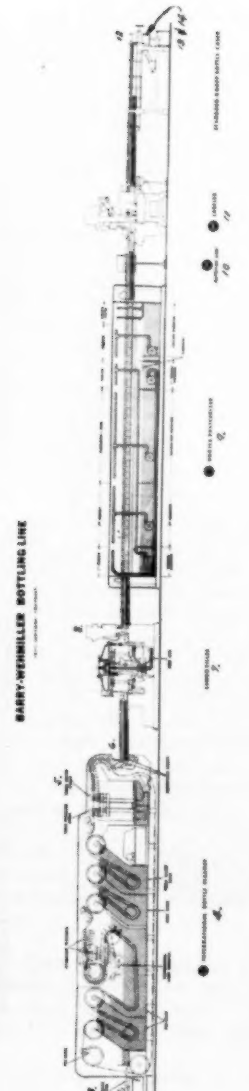
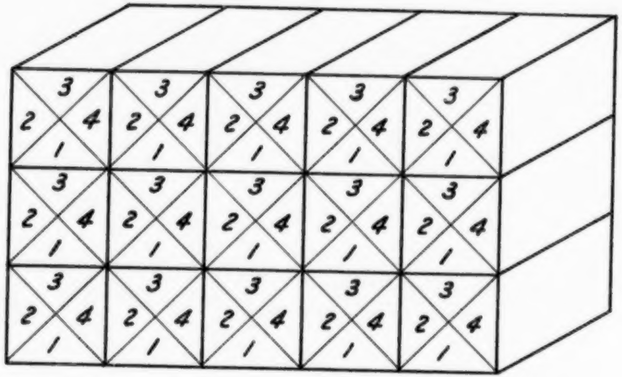


FIG. 2



EACH BOTTLING LINE CONSISTS OF 15 SEPARATE UNITS

EACH UNIT HAS 4 POSSIBLE SOURCES OF DEFECTS

1. MECHANICAL
2. ELECTRICAL
3. SUPPLIES & SERVICES
4. PERSONNEL

The following slide (Figure 4A) indicates the form used for each inspection tour and has been filled in with values typical of such a tour. For example, on this particular inspection there was one Class I mechanical defect, no electrical defects; there was one Class I personnel defect, and there were no Class I supplies and services defects. Among the Class II defects there was one mechanical, one electrical, no personnel or services and supplies. The Class III defects found were, no mechanical, one electrical, one personnel, and one services and supplies defect. Totaling up the scored demerits (wd) for each class, the sum would be forty-nine for this particular tour, and the demerits per unit would be 3.26.

In the work presented here a control chart was constructed after thirty hypothetical inspection tours had been made. This next slide (Figure 4B) will indicate the method used in the construction of the control chart. As can be noted, "w" equals the assigned weight of the class of defect; "d" is the number of defects observed in this base period for each class of defect; "n" is the number of units in a sample, which as mentioned earlier is 15; "N_T" is the total number of units in the base period or 15 X 30; "Du" is the average demerits per unit; and "C_s" is the constant of variance for this particular series of inspections. The calculation of the control limits for this operational period was then found to be: 1.325 for the 3σ upper limit, .984 for the 2σ upper limit, and .322 for the average.

Control charts for the period of study would appear as follows. (Figure 5A & 5B) Figure 5A indicates a chart based on the demerits per bottling line, and 5B is based on the defects per bottling line. In Figure 5A each point is the average demerits per unit for the inspection tour. Sample number 20 has fallen outside of the 2 limits and is very nearly out of the 3 limits of the chart and calls for an investigation. However, we can afford to run some risk on this sample but should be on guard for a recurrence of such a combination of defects.

It is interesting to note the chart in Figure 5B, which is based on the same thirty inspection tours but recorded on the basis of defects per unit rather than demerits per unit, shows that there would have been no warning signal when it was needed but might have aroused us unnecessarily at sample 8 where eight minor defects were reported.

In installing a system of this type all operators should be fully advised of the demerit weight of each type of demerit. Each bottling line is scored weekly and comparisons are made on the operating efficiency level of each line.

It should be noted that the assumption has been made that the number of defects are independent variables and are subject to the variations in the inspection tours; also that the ratio of the number defects to the possible number of defects is small so that we can assume d = d .

When using a weighted defect system as discussed here, there are a number of beneficial results that can be realized. By listing and weighting the quality level of the bottling operation the weak points of the operation can soon be discovered and corrective action can be aimed at these weak points. Follow-up of such a program results in better trained

Fig. 4A

QUALITY INSPECTION REPORT

Inspector ______________________ Sample No. ________

Unit No. ____________ Line Foreman __________ Date ________

Class #	W	Mech.	Elec.	Personnel	S & S	d	wd
I	10	1		1	1	3	30
II	5	1	1			2	10
III	3		1	1	1	3	9
					Total Demerits		49
					Total Demerits/15		3.26

Fig. 4B

QUALITY LEVEL CONTROL LIMITS

Based on 30 Inspection Tours

Class No.	Weight w	# of defects d	wd	w^2	w^2d	d/n_r
I	10	17	170	100	1700	.037
II	5	41	205	25	1025	.091
III	3	59	177	9	531	.128
			552		3256	

w = assigned weight

d = number of defects observed

n = number of units in sample

n_r = number of units in base period $30 \times 15 = 450$

Du = average demerits per unit = $wd/n_r = 552/450 = 1.22$

C_s = constant of variance = $w^2d/n_r = 3256/450 = 7.2$

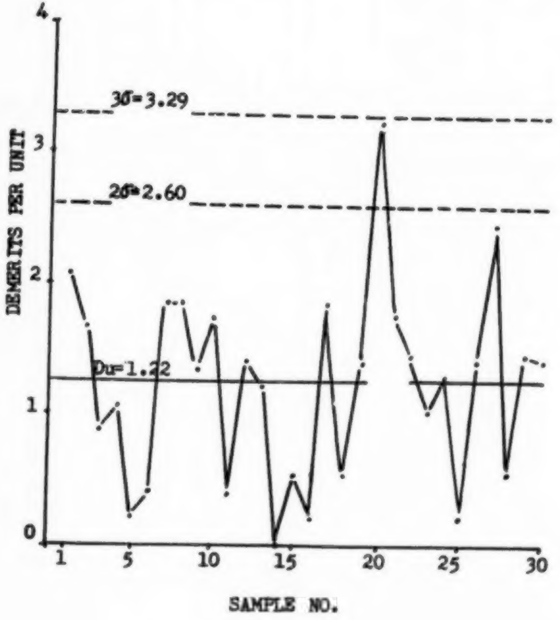
$\sigma Du = \sqrt{C_s/n}$ $\quad$ $Du \pm 3\sqrt{C_s/n}$ $\quad$ $Du \pm 2\sqrt{C_s/n}$

$\pm 3\sqrt{.48}$ $\quad$ $\pm 2\sqrt{.48}$

3σ - Upper Limit = 3.296 $\quad$ 2σ - Upper Limit = 2.604

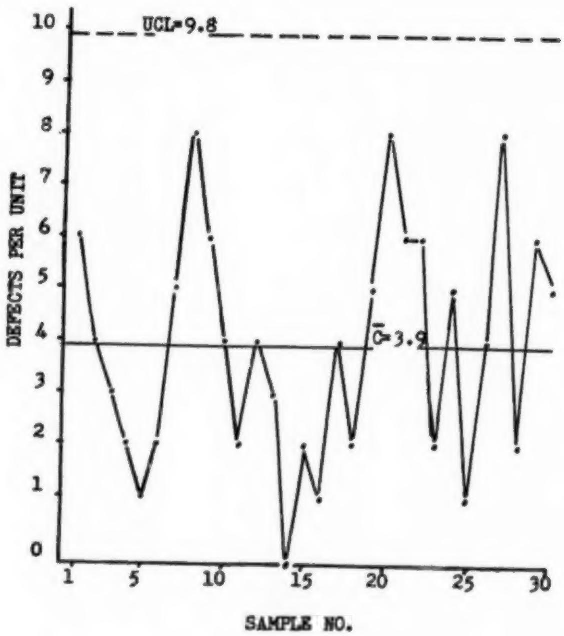
3σ - Lower Limit = 0 $\quad$ 2σ - Lower Limit = 0

FIG. 5A



DEMERITS PER UNIT
SAMPLE NO.
3σ=3.29
2σ=2.60
Du=1.22

FIG. 5B



DEFECTS PER UNIT
SAMPLE NO.
UCL=9.8
C̄=3.9

operators, better maintenance of mechanical and electrical functions, better services and supplies, more efficient methods and handling, and an increased responsibility of each man working in the bottling unit for producing and maintaining a high quality of line operation.

With the increased production rates of today, and with an ever growing demand for faster and faster production lines, it becomes evident that examination of single components of the bottling line will never reveal the true operating level of the entire unit.

It has therefore been the purpose of this paper to examine a method of inspecting the entire bottling operation as an indivisible unit. This program may seem ambitious and may not be suitable to every bottling department; however, the general theory of this method of inspection with variations dictated by the operation involved has, in our opinion, many beneficial possibilities.

References:

H. F. Dodge - "A Method of Rating Manufactured Products", Bell Telephone Laboratories, Reprint B315.

Electrical Engineering, March 1946 - "Statistical Methods in Quality Control"; No.X - "Classification of Defects and Quality Rating".

QUALITY CONTROL, INDUSTRIAL ENGINEERING, AND OPERATIONS RESEARCH

Warren E. Alberts
United Air Lines

Thomas Carlyle, the philospher, once said, "The purpose of education is not knowledge, but action." The purpose of this paper is not to increase your knowledge, but to excite you to action. It is not a discourse on techniques, but a challenge -- a challenge to men in the Industrial Engineering and Quality Control fields to meet management's needs, not tomorrow but today.

Reading some of today's literature, one would gather that an Industrial Engineer is a rather dull fellow who can manipulate a stop watch and is constantly taking time studies and making process charts. Further, he is purely a shop man and is unaware of such things as interdepartmental coordination, sales quotas, human aspects, organizational theory, policy decisions, and, of course, is never allowed to look at operating statements and only speaks to department heads when he receives his ten-year pin.

On the other hand, I often hear a Quality Control Engineer spoken of as an ex-inspector who has somehow become exposed to statistical theory and is constantly running around posting and looking at control charts, all the time muttering that no one appreciates his work and that if only the production and engineering departments would listen to him and not look at him as an inspector in sheep's clothing he could immediately set the operation right.

You and I know that these concepts are untrue, but don't kid yourself that they are not real impressions and must be overcome by demonstrated proof to the contrary. It is natural to resent the narrow scope attributed at times to Quality Control and Industrial Engineering, but one can't fail to be aware of the reasons. In the first place, many company managements don't realize that the approach and professional skills of the Industrial and Quality Control Engineer are applicable to almost all management problems and are not confined to shop production and inspection functions. This is not altogether their fault because much of the work being done, the nomenclature, trade articles and our very titles lead them to the conclusion that we are primarily concerned with shop processes.

Secondly, too many men in the field have specialized in the repetitive use of certain techniques and concentrated on refining small pieces of the over-all industrial or business process. No wonder management gets to thinking of them as methods men, time study men, chart men, or quality inspectors. Those Industrial or Quality Control Engineers who do visualize the broader use of their tools, and there are quite a few, often contend they can't get to use them due to managements' restrictions on their scope, their organizational status, etc. Sounds like a vicious circle, doesn't it? Well it is until, in a given situation, something breaks it.

The recent growth of the Operations Research or systems engineering concept is timely because it is going to help break that circle and jar many an Industrial Engineer, Quality Control Engineer, and executive from the comfortable ruts which they have shaped for themselves.

Operations Research or whatever you want to call it is going to represent different things to each individual and company depending on how they have used and combined the various concepts and tools of scientific management. From our experience with it I think of OR as one part management engineering, one part statistical quality control, a drop of higher mathematics, mixed well by a team of men operating on a problem with their heads in the clouds, their feet on the ground, and no holds barred. Oh, I almost forgot - plus lots of time and a healthy budget.

My first experience with Operations Research was in 1944 when, as Director of Operations for the Second Air Division, an Operations Analysis Group was attached to our headquarters, headed by a Doctor of Mathematics from Harvard. The doctor did not bring an electronic computer with him; only a piece of paper, a pencil and a slide rule. He did not use any fancy mathematics but got to work correlating the bombing accuracy of our groups to the formation which they were flying at time of release. Also, of extreme interest to us at the time was his calculation of the chances of being hit by enemy fighters in a particular type of formation.

It is pointless to argue what is new about Operations Research or whether it is merely Industrial Engineering or Quality Control dressed up in a new suit and digging deeper into bigger problems. What is important is your taking whatever is new in the concept to you as an individual and using it to broaden your outlook, give new meaning to your skills, and excite you to add new tools to your kit. The complexity of managements' problems today requires the best that Quality Control, Industrial, or Research Engineers can contribute. It has already been conceded that probability and statistical theory are the most important single tools of OR, so as far as skills go, you gentlemen are in on the ground floor. A word of caution here: the broader a problem, the less important becomes the tool and the more important the attitude, imagination, experience, and ability of the individual involved.

Don't be thrown off base by the technical literature on OR and the implication that it is reserved for a few. There are a lot of theories, loads of techniques, hundreds of formulas, but except for military applications and a few classics in industry, the ground is virtually unplowed and the men who are developing these techniques are patiently waiting for some one to apply them and prove their worth.

So don't worry about what your present title is or what certain types of work are supposedly called. If you catch the spirit - go to it - and let the results speak for themselves.

Although management's need for these new concepts and broader use of tools is very real, don't expect your executives to suddenly tap you on the shoulder and say you're just what the doctor ordered; you've got to sell and prove their worth.

Reflect that management, too, has become concerned over the rapidly diminishing returns resulting from the continual refinement and adjustment of their existing work processes, organization, and equipment. They are beginning to sense that changing attitudes, needs, and hardware call for periodic overhauls of the whole process. The com-

plexity of such a task and its implications are causing them to gradually realize the need for ways of measuring and understanding the true nature of their entire operation.

I say "true nature" because many executives are becoming aware of the fact that of all the reports, statistics, and financial figures they see, very few are designed to tell them in a timely way precisely what is happening and why. They are beginning to realize that very little material is gathered to enable management to manage, most of it is to tabulate what different functions of the company believe to be a measure of results.

Also, the increasing size, complexity, and cost of doing business too is placing a bigger premium on each major decision and at those times the lack of understanding, facts, and over-all measurements make them realize how little they have on which to base their judgment. In extremely critical decisions of a long range nature, they dream of a way to test the alternatives and in some way forecast the varying probability of success.

The appearance of high-speed computers and data processing equipment on the scene has raised their hopes for getting the right information on time and being able to release the energies of a large percentage of their personnel for higher skilled work. The fact that one out of every five workers in the United States is engaged in paper work is of serious concern.

It looks like a set up, doesn't it? Well don't be fooled. Management, like you, have to be sold and shown by examples of how most of their needs can be met by the imaginative application of the very tools in our kits.

To me, an Industrial Engineer and Quality Control Engineer have a lot of knowledge and experience to exchange. More Industrial Engineers must become acquainted with the natural laws of variation, probability and their infinite application. Likewise, Quality Control Engineers should learn the all-important scientific approach to special problems and develop a feel for methods, organization and the problems of human beings at work. Maybe I am suggesting a coalition; it makes little difference. Ultimately, it all will depend on the individual and his interests and abilities. You can be sure there will be enough work for both.

To give you an idea of how we in United are utilizing the varying approaches and techniques of Industrial and Quality Control Engineers, let me explain our set up.

First of all, the Industrial Engineering Department in United reports to a non-operating administration headed by a Vice President - the Vice President of Economic Controls - who also has as his responsibility such things as market research, economic forecasts, airplane schedules, and cost control. We in Industrial Engineering provide an internal consulting service to the rest of the company in all fields of management engineering. The little card which we proudly issue to our prospective clients sets forth the following:

1. We specialize in isolating and defining problems and offering

specific solutions for your decision.

2. Time, objectivity, a trained approach, and all the tools of scientific management are at your service.

3. We respect our clients' confidences and carefully consider the human aspects of each problem.

4. Services available on request but subject to current commitments to other clients and programs.

Our clients include the President, members of the General Staff, department, division and section heads. Some of the services which we offer are as follows:

1. Management Studies
2. Special Surveys
3. Organization Studies
4. Methods & Work Simplification
5. Work Standards
6. Quality Control
7. Operations Research
8. Planning & Programing
9. Facility Planning
10. Job Evaluation
11. Job Analysis
12. Form Design
13. Regulations
14. Incentives

Now let's see how we are organized to do our job. There are five groups: Organization Planning, Work Analysis, Regulations & Forms, Quality Control, and Operations Research. Under Work Analysis we have facility planning, standards, methods, and special projects. We do not worry too much about formal organizational lines but, depending on the project, bring together those men who have the skill to best attack the problem at hand. Although we must serve over 16,000 people in the company, our goal is not to see how big we can get but how good a service we can render. Our aim is to provide the company with those skills and tools which it would be unable to get in any other way; to provide an impartial and objective viewpoint when necessary, and to introduce and train other people in the use of new management tools. We make a conscious effort to multiply ourselves by continual training programs in methods, quality control, work simplification, etc. Other than our standards setting job, if any phase of our work gets repetitive or routine, we find a way to turn it over to an operating administration.

At present the functions of Quality Control and Operations Research are separated; however, the time may come when Quality Control and Operations Research might well be combined into a group known as Systems Engineering or Applied Statistics.

As a result of our Organization Planning responsibilities, we have been able to establish Quality Control groups in our Operating Administrations and assist them in developing programs which they administer. This leaves Mr. Dalleck, our Staff Superintendent of Quality Control, who is responsible for sparking our company-wide Quality Control Program, free to develop new applications and techniques, and to train staff and supervisory personnel.

We firmly believe that the more people we expose to the basic theories of probability, sampling and the natural laws of variation, the more productive and understanding they will become in their own tasks. This process takes time because the average accountant's, engineer's,

and supervisor's background and education rest pretty firmly on empirical concepts and the use of averages and fixed values.

To illustrate the scope of our Quality Control activities and how we try to apply statistical techniques wherever we can, here are a few examples of some of the projects that are or have been worked on:

1. A work sampling check on a large shop in our Maintenance Base to determine the reasons for a low utilization on standards.

2. A sampling of supervisory activities in the Accounting Department to determine distribution of effort.

3. A six months' experiment between three major airlines testing the use of sampling techniques to settle interline accounts. Mr. Dalleck gave a paper on this at the ASQC Convention in 1954. We feel this and similar applications are going to start a revolution in the accounting field because many are beginning to realize that "inductive accounting," as one man called it, is often more accurate than 100% verification.

In July of 1954 we decided to undertake a full blown Operations Research approach to our system aircraft routing and maintenance problems. Because we felt we were basically familiar with many aspects of the approach and had never been restricted in tackling over-all company problems regardless of where they led, we decided to strike off on our own. This only was done, however, after we had talked to several consulting firms and research organizations, and had scanned reports of OR work. Since then, in talking to various people at computing and research centers, we find that we have a major project by the tail.

The project basically concerns itself with the whole process of providing serviceable aircraft to meet schedules over our 13,250 mile system. This naturally leads into performance, location of facilities, manpower scheduling, maintenance plans, flight delays, etc. Our objective is to determine the true nature of our existing operation and then provide management with station and system models on which to test their ideas and plans.

As members of the team, we picked a top notch aerodynamicist with a good background in higher mathematics from our Engineering Department in San Francisco and a 25-year veteran in the flight operations field from the Flight Dispatch Manager's group at our Denver Operating Base. Recognizing the importance of statistical techniques, we assigned our Staff Superintendent-Quality Control and, in addition, an Industrial Engineer who had worked on procedural and systems problems.

After nine months we all agree with Morse and Kimball that the most important single mathematical tool of Operations Research is probability and statistical theory. Little has been written on the problems of a team approach to a really complex business problem, but we feel as though we already have enough material for a book. Much of the work is routinely time consuming, involving decisions as to what material is needed, how it can be collected, collecting it, deciding on how it will be analyzed, making test analyses, and problems of hand vs. machine computation.

Always there is the need to balance the varying viewpoints of the team members, to keep the project from digging too deeply or passing too lightly over a critical factor. Getting the material you need and making sure it is without bias is quite a problem in itself.

Enough about the OR project - let's take a look at where we are. We've talked about how other people look at the Industrial and Quality Control Engineer, the need for and application of our skills to higher level problems, the impact of OR, and how one company is trying to approach its opportunities in these fields. Unless I've missed the beam, your chest should be slightly inflated and you should have a new perspective regarding your skills and your future.

Now this feeling won't last unless you do something about it. Remember, I said the purpose of this paper was not to increase your knowledge, but to spark you to action.

You probably are way ahead of me as to just what you're going to do, but maybe this list of do's and don'ts will prove of value.

1. Don't let titles or the nomenclature of the day narrow your vision as to the job that you can do or that which needs doing. Quality Control has been dancing with inspection and Industrial Engineering with production for so long they both have almost missed some very attractive partners.

2. Do try and grasp the impact of your statistical and probability skills - their application to almost every problem. But don't fall for the fatal fascination of techniques as such - they mean nothing unless they are imaginatively put to work.

3. Don't be confused by the mathematical jargon of Operations Research or Systems Engineering - you have their basic tools in your kit. Concentrate on assimilating what's new in concept and approach.

4. Do undertake an application of your skills to a problem off the beaten path - no matter how small - even on your own time or for someone in another department. If it's good, let them have the credit. Soon you'll be getting requests and recognition of your skills. Then is the time to submit your proposal for an attack on a bigger management problem and you're on your way.

5. Do try to spread the word - exchange ideas and experiences.

Here is your challenge and your opportunity! Everything in the universe is subject to the natural laws of variation. You who have the tools of measurement and prediction must demonstrate their application to management problems and other fields of endeavor with imagination and understanding.

DEPARTURES FROM RANDOMNESS

Frank G. Norris
Wheeling Steel Corporation

"What we do not see, we tread
upon, and never think of it."

The distribution of defects upon the surface of a metallic product differs from the distribution of defectives among many pieces of a batch or lot. In order to help to clarify various aspects of this difference three questions are suggested for consideration by the panel.

Figures 1 to 6 show distributions of defects (attributes) over an area. For convenience in presentation the area is square (100 x 100). Much the same reasoning would apply if the area were rectangular. It could be the curved surface of a pipe, or the almost linear surface of a wire.

Figures 7 and 8 show two sampling plans for selecting ten areas each containing one percent of the total area. Each circle (or square of equivalent area) will be called a unit sampling area.

In each of the six tables the observed distribution of defects is compared with the results of using each of the two sampling plans.

In the first column of each table is the number of defects. In the second column is the number of unit areas expected to contain a given number of defects based on the assumption of the Poisson distribution. In the third column is the actual (observed) number of unit areas with a given number of defects. In the fourth column are the results of sampling according to the sampling plan of Figure 7 (regular pattern). In the fifth column are the results of sampling according to the sampling plan of Figure 8 (random).

1. IS THERE A NEED TO DISTINGUISH DIFFERENCES IN THE PATTERN OF THE DISTRIBUTION OF DEFECTS ON THE SURFACE OF A METAL PRODUCT? i.e. DIFFERENCES SUCH AS ARE ILLUSTRATED AMONG FIGURES 1, 2, 3, 4, 5, 6.

This question must be answered by the consumer because the use of the product determines the answer.

The purpose of this panel is to direct attention to this question rather than to answer it.

If the material is to be used to supply a large number of small blanks, each of which must be perfect, the distribution of Figure 1 is much worse than that of Figure 6.

If the surface is to be covered with paint, or enamel, or insulating material, Figure 1, 2, or 3, or even 4 which contains a greater number of defects, might be more acceptable than Figure 6.

2. HOW CAN DIFFERENCES IN THE PATTERN OF THE DISTRIBUTION OF DEFECTS BE DESCRIBED - OR SPECIFIED?

Assuming that the pattern does make a difference, the next problem is description.

The classification random and not random is not sufficiently definitive. There is a limited number of possible distributions of 100 defects on a grid of 10,000 locations. Some fraction of these can be considered random distributions. One of these distributions (believed to be random) is shown in Figure 3. Each of the other distributions of 100 defects is non-random, but they are not the same.

Figure 1 could be modified slightly without making any material difference. When does a modification such as Figure 2 become great enough that it is considered random?

How can special patterns such as stringers or clusters be defined?

What other types of patterns should be considered?

3. HOW CAN DIFFERENCES IN THE PATTERN OF THE DISTRIBUTION OF DEFECTS BE MEASURED?

The problem of measurement may have to be answered before considering the previous question of description.

Two sampling plans are illustrated in Figure 7 and 8. Figure 7 is a regular pattern, a slight modification of the method of sampling edge, center and edge of a sheet or strip.

Figure 8 is one of many possible random samples.

Either sampling plan is adequate to distinguish Figure 4 and Figure 3. i.e. increase in the number of randomly spaced defects.

When the pattern of the distribution changes, how should the selection or interpretation of the sample be changed to distinguish Figure 7 from either Figure 4 (i.e. a larger number of defects) or from Figure 1 or 3 (the same number of defects arranged in a different pattern)?

INDEX TO FIGURES

TABLE 1 - 100 DEFECTS UNIFORMLY SPACED

Number of Defects per Unit Sampling Area	Frequency Expected (Poisson Law)	Frequency Observed	Frequency in 10 Samples Regular Spacing (Figure 7)	Frequency in 10 Random Samples (Figure 8)
0	36.8	0	10	0
1	36.8	100	0	8
2	18.4	0	0	2

TABLE 2 - 100 DEFECTS RANDOMLY LOCATED WITHIN UNIFORMLY SPACED AREAS

Number of Defects per Unit Sampling Area	Frequency Expected (Poisson Law)	Frequency Observed	Frequency in 10 Samples Regular Spacing (Figure 7)	Frequency in 10 Random Samples (Figure 8)
0	36.8	0	4	2
1	36.8	100	3	4
2	18.4	0	3	4

TABLE 3 - 100 DEFECTS RANDOMLY SPACED

Number of Defects per Unit Sampling Area	Frequency Expected (Poisson Law)	Frequency Observed	Frequency in 10 Samples Regular Spacing (Figure 7)	Frequency in 10 Random Samples (Figure 8)
0	36.8	37	5	3
1	36.8	38	2	6
2	18.4	17	2	1
3	6.1	6	1	0
4	1.5	0	0	0
5	.3	2	0	0
6	.1	0	0	0

TABLE 4 - 200 DEFECTS RANDOMLY SPACED

Number of Defects per Unit Sampling Area	Frequency Expected (Poisson Law)	Frequency Observed	Frequency in 10 Samples Regular Spacing (Figure 7)	Frequency in 10 Random Samples (Figure 8)
0	13.5	12	1	1
1	27.1	27	4	1
2	27.1	28	2	3
3	18.0	21	2	5
4	9.0	8	0	0
5	3.6	3	1	0
6	1.2	1	0	0
7	.4	0	0	0
8	.1	0	0	0

TABLE 5 - 100 DEFECTS LINEAR PATTERN

Number of Defects per Unit Sampling Area	Frequency Expected (Poisson Law)	Frequency Observed	Frequency in 10 Samples Regular Spacing (Figure 7)	Frequency in 10 Random Samples (Figure 8)
0	36.8	52	4	5
1	36.8	13	2	2
2	18.4	20	2	2
3	6.1	13	2	1
4	1.5	2	0	0

TABLE 6 - 100 DEFECTS CLUSTER PATTERN

Number of Defects per Unit Sampling Area	Frequency Expected (Poisson Law)	Frequency Observed	Frequency in 10 Samples Regular Spacing (Figure 7)	Frequency in 10 Random Samples (Figure 8)
0	36.8	83	7	10
1	36.8	2	2	0
2	18.4	2	0	0
3	6.1	5	0	0
4	1.5	0	0	0
5	.3	2	0	0
6	.1	0	0	0
7		2	0	0
8		1	0	0
9		1	0	0
16		1	0	0
17		0	1	0
22		1	0	0

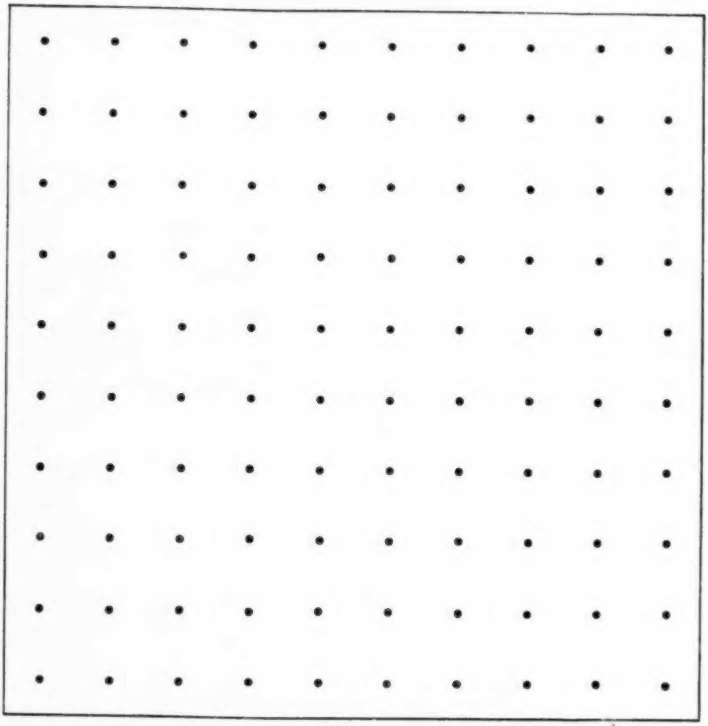


Fig. 1 100 DEFECTS UNIFORMLY SPACED

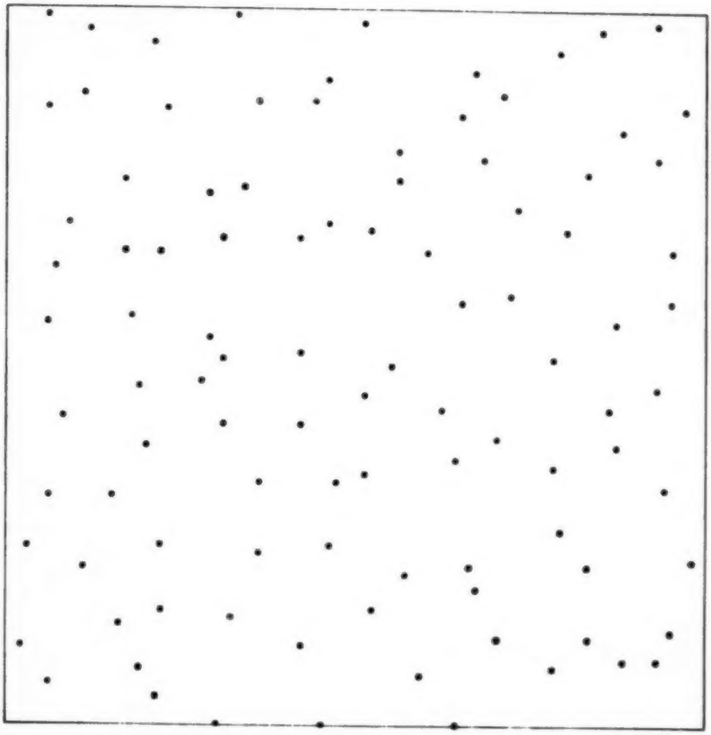


Fig. 2 100 DEFECTS RANDOMLY LOCATED WITHIN UNIFORMLY SPACED AREAS

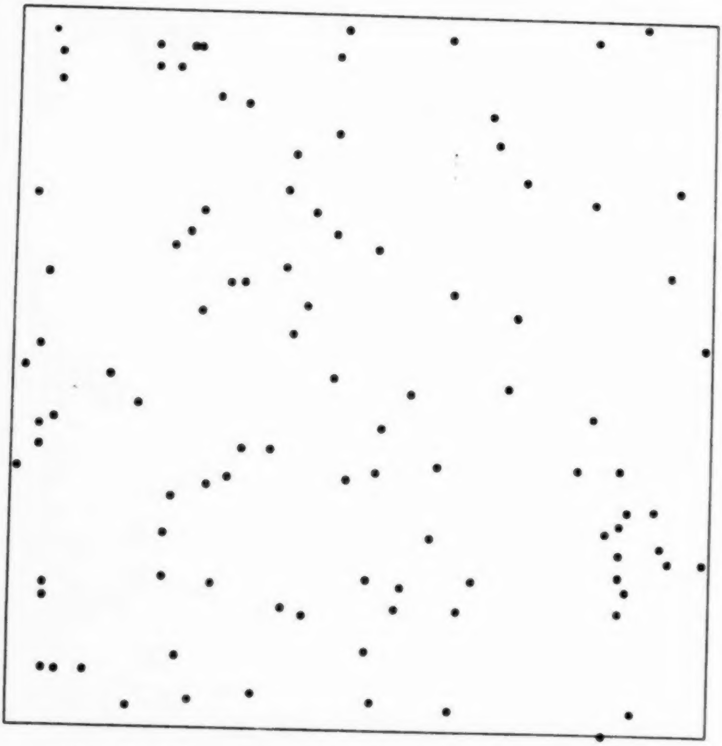


Fig. 3 100 DEFECTS RANDOMLY SPACED

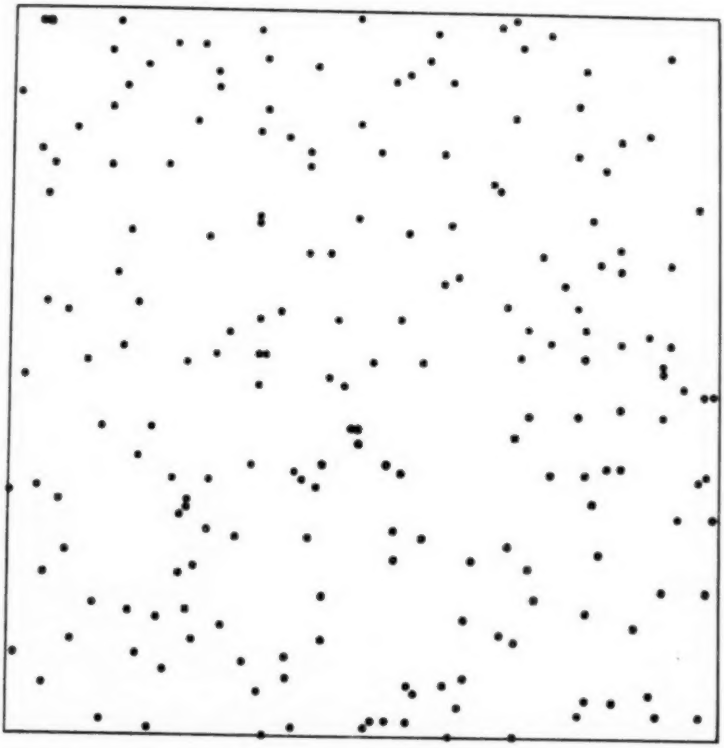


Fig. 4 200 DEFECTS RANDOMLY SPACED

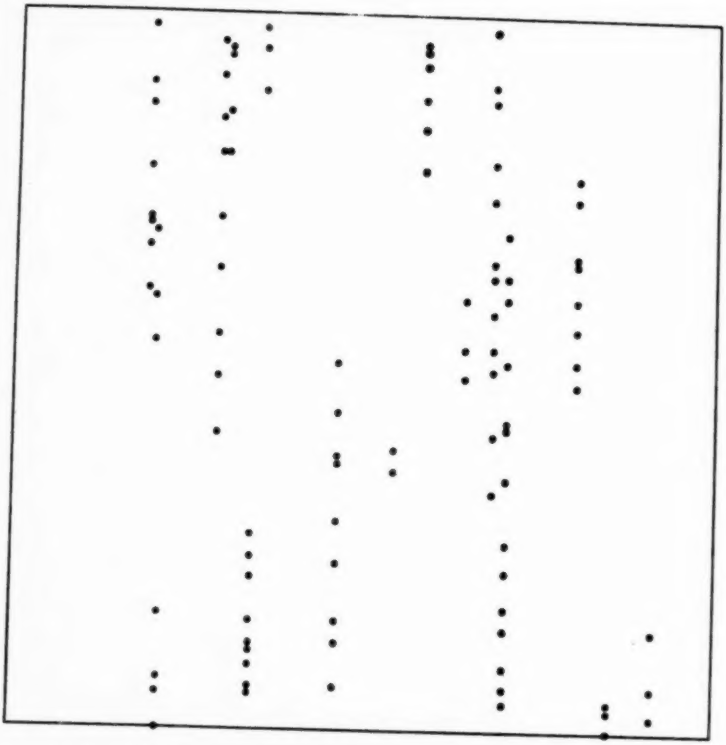


Fig. 5 100 DEFECTS LINEAR PATTERN

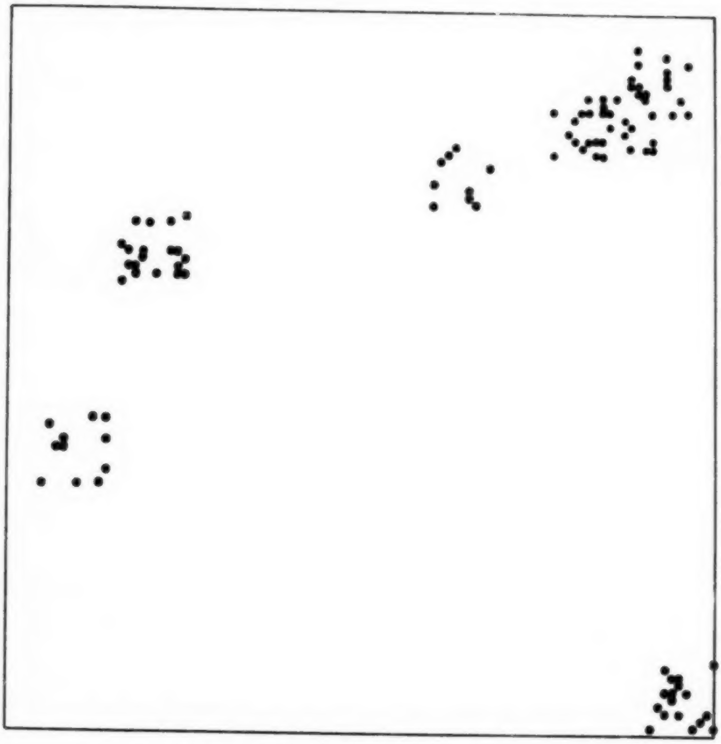


Fig. 6 100 DEFECTS CLUSTER PATTERN

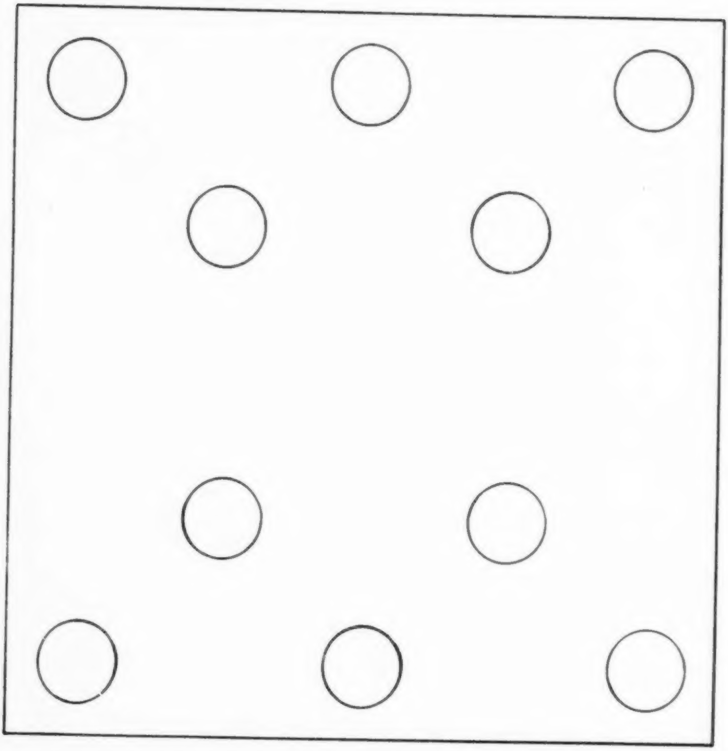


Fig. 7 SAMPLING PLAN 10% OF AREA
REGULAR SPACING

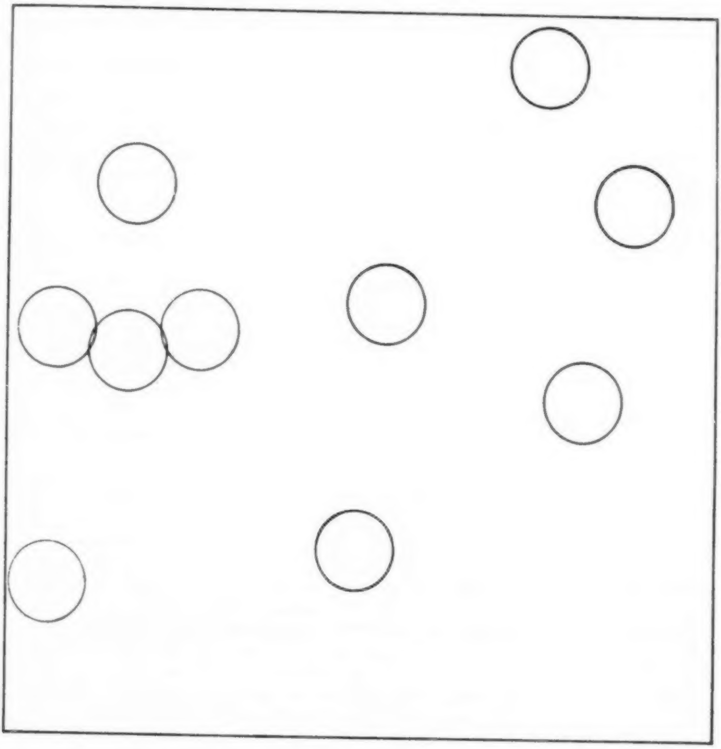


Fig. 8 SAMPLING PLAN 10% OF AREA
RANDOM LOCATION OF SAMPLES

PRACTICAL LINEAR PROGRAMMING APPLICATIONS

Harry T. Schwan
Methods Engineering Council

One of the brightest spots on management's horizon today is a new tool called Linear Programming. This new tool is already helping management to make better decisions on some of its most complicated problems. If the applications that have already been made are an indication of its usefulness, then I am sure it is going to be one of the most valuable aids to management decision-making which has turned up in the past fifty years.

Linear Programming is useful because it adds precision to the process of decision-making. It permits management to move along a positive course of action knowing that that course of action is the best it can do under its own present circumstances. It provides management with facts which are predicated upon a consideration of the total problem rather than piecemeal consideration of various parts of the problem.

Areas of Application

Linear Programming has already been applied with excellent and even spectacular results to determine:

1. The most profitable manufacturing program.
2. The best inventory strategies.
3. The effect of changes in purchasing and selling price.
4. Whether to make or buy certain component parts.
5. The most profitable product mix.
6. The best location of plants.
7. The best location of warehouses and distribution outlets.
8. The lowest cost machine or manufacturing schedule.

This is only a partial list of applications, but it does point out the type of problem upon which Linear Programming is most useful. I'm sure you noticed that there are some common characteristics in each one of these problems.

First, to make a decision on any one of these problems the manager must consider a very large number of factors. Second, these factors are almost always inter-dependent so that the manager must consider them both individually and in relation to each other. And, third, the manager is faced with having to choose <u>one</u> solution or course of action from among several obvious courses of action and, perhaps, several others which are not so obvious.

I'm sure you will agree with me that it would take an unbelievably brilliant human mind to understand, weigh, balance, and keep in their proper perspective all of these factors, and to pick the best solution. In the past, most of us who have had to face these situations have done a lot of decision-making based on experience, feel, intuition, hope, and pure "guesstimate." We have had to do this because we had no other tool that could add precision to our deliberations. The best human minds have not been capable of accommodating, with precision, problems which are this complicated.

How Linear Programming Assists Management

Linear Programming, from the manager's point of view, is both an <u>approach</u> to the formulation and statement of these complicated problems

and a set of mathematical procedures that enables him to handle the complications and select the best course of action. Stated another way, Linear Programming helps the manager to:

1. Organize the facts and information about a problem.
2. Analyze all possible alternative solutions to the problem.
3. Select the best course to follow under his own conditions and limitations.
4. Plan the specific steps required to get the best results.
5. Re-evaluate the plan when conditions change.

Let's consider each of these points for a moment.

We have found that, when we go to organize a problem so that it can be handled by the Linear Programming mathematical procedures, we nearly always gain a new perspective and keener insight into the problem. In some cases this added clarity has given us a different and more valuable picture of the true problem. It has led us more surely to causes rather than effects. To permanent solutions rather than stop-gap expedients.

We want to analyze all reasonable courses of action, because, for various reasons, we may not choose to follow the best course of action. With Linear Programming we can readily determine what we are giving up by following a course of action other than the best one.

Linear Programming identifies the best course of action for us to follow under the conditions which affect us. We can build into our solution a true reflection of our own limitations or restrictions in marketing, production facilities, finance, manpower, and many other practical operating factors.

The solution we obtain comes out in specific, quantitative terms -- how many shall we make -- how shall we make them -- where shall we make them. With this type of information we can plan the specific actions we will need to take to get the results that are possible.

Finally, and this is exceedingly important, we can re-evaluate our plans and programs when conditions change or when we think they might change. This means we have a before-the-fact tool as well as an after-the-fact tool. It means that we have a tool that is practical in the ever-moving, ever-changing, business atmosphere in which we must make decisions.

Linear Programming in Action

The best way for you to get an idea of the way Linear Programming works is to follow me through a typical problem. This problem illustrates how Linear Programming can be helpful in planning for maximum profits. Figure 1 illustrates the situation that confronts us.

We have a plant that is capable of making two products, A and B. To make these products we have available certain facilities. The nature of the products is such that two operations are required on each, but the first operation must be performed on Machine Group 1. In Figure 1 this Machine Group is designated M1. For operation 2, however, we have three choices as to where it can be performed for either product. We can use Machine Group 2 (M2 in Figure 1) on straight time, Machine Group 2 on overtime (M2A in Figure 1), or we can use Machine Group 3 (M3 in Figure 1).

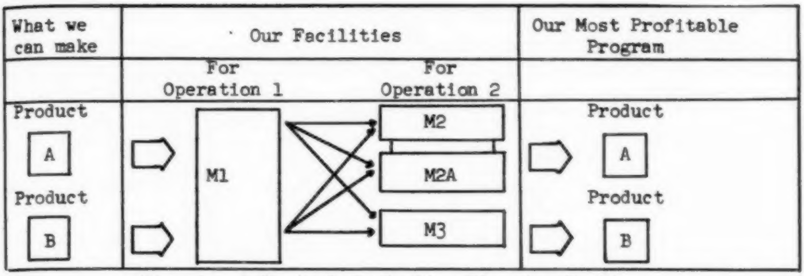


Fig. 1 - A Typical Production Planning Problem

Considering what we can make and what we have to make it with, our management has asked us what manufacturing program will provide the most profitable use of these facilities. They want us to tell them how many of each product to make, which of the three manufacturing alternatives to use, and how much profit this program will produce.

Before we can determine our best course of action, we need the additional information given in Figure 2.

Operation	Mach. Group	Hours per 1000 Pieces						Hours Available
		Product A			Product B			
1	M1	2	2	2	5	5	5	1000
2	M2	3			8			600
2	M2A		3			8		200
2	M3			4			10	800
Profit/piece		.85	.60	.70	1.60	1.40	1.30	

Fig. 2 - Manufacturing Information on the Facilities Given in Fig. 1

Naturally, there must be a limitation on how much time can be spent to produce our two products. These time limitations are given in the right-hand column. We have 1000 hours available on M1, 600 hours available on M2, 200 hours of overtime available on M2A, and 800 hours available on M3.

The data in the center of the chart show the time in hours to manufacture 1000 pieces for each operation, on each product, and by each manufacturing alternative.

On the bottom line you see some very important figures which give us the unit profit for each product when it is manufactured by each manufacturing alternative.

Before we look at the answer to this problem, let's remember that our management has asked us for the most profitable <u>manufacturing</u> program, and that no restriction has been placed on our ability to sell whatever we decide to turn out. We will build in the sales restriction later.

The result of applying Linear Programming to this problem is given in Figure 3.

Production Quantities		Facilities Used	Profit
Product A	Product B		
200,000		M1 & M2	$170,000
66,667		M1 & M2A	40,000
200,000		M1 & M3	140,000
466,667	None		$350,000
Total			Total

Machine	Hours Needed			Total Hrs. Needed	Hours Avail.	W
M1	400	133.3	400	933.3	1000	0
M2	600			600	600	283
M2A		200		200	200	200
M3			800	800	800	175

Fig. 3 - The Most Profitable Production Program as Determined Through Linear Programming

The largest possible profit is obtained when all machines are used to make Product A. A lower profit will be made if any of Product B is produced despite the fact that the profit per piece on Product B is much higher than on A. As Figure 3 indicates, the program which should be followed is:

1. Use M1 and M2 to make 200,000 pieces of Product A which will give $170,000 profit.
2. Use M1 and M2A (M2 on overtime) to make 66,667 pieces of Product A which will give $40,000 profit.
3. Use M1 and M3 to make 200,000 pieces of Product A which will give $140,000 profit.
4. The maximum profit, is, therefore, $350,000.

The time required of each Machine Group to follow this program is also given in Figure 3.

The first interesting thing you should note from the table is that you should use less than the total capacity of M1 in order to obtain maximum profit. If you fall into the trap of using all of M1's 1000 hours, you will lower your profit. You will simply succeed in building up your work-in-process inventory.

Next, look at the W column. The $283 figure opposite M2 means that you can increase your profit by $283 for each hour of additional capacity you can provide on M2. You will find it particularly interesting to note that you can make more profit from an extra hour of overtime on M2 than from an extra hour of straight time on M3. These figures hold, of course, only until you have used all of the capacity available on M1.

So you see that you have definite and specific answers as to what to do and how to do it for maximum profits. The total time figures provide you with a basis for planning manpower and maintenance activities. The unit quantity and time figures provide you with firm figures upon which to base purchasing, inventory, and sales plans. You can use these figures knowing that you have looked at your total problem and that all of your planning and action will be coordinated and aimed at maximum profit.

The Effect of Introducing Added Restrictions

Let us look, now, at what our program might be with an additional

restriction. Suppose that 100,000 pieces of Produce B have been sold and must be produced. All other conditions remain the same. The most profitable production program becomes that shown in Figure 4.

Product	Manufacturing Alternative			Total
	M1-M2	M1-M2A	M1-M3	Production
A	200,000	0	12,500	212,500
B	0	25,000	75,000	100,000
	Maximum Profit = $311,250 Sacrifice Profit = $ 38,750 Sacrifice Production = 254,167 units of A			

Fig. 4 - Most Profitable Program if You Must Make 100,000 Units of Product B.

We see that our facilities are now used in a different way. Most of Product A are produced on Machine Group M1 in conjunction with M2. Only a small amount is produced on Machine Group M3, and M2A is not used at all. Three quarters of the production of Product B is obtained from using a combination of Machine Groups M1 and M3 with the balance coming from M1 and M2A.

The profit for this program is $311,250. By comparing this to the original profit of $350,000, you can see that by producing 100,000 pieces of B, $38,750 of profit are foregone. In addition, the number of Product A that will be available for delivery is reduced by 254,167.

The questions now raised are:

1. "Are the customers to whom these 100,000 units of B were committed worth a sacrifice of $38,750 in profit," and
2. "If we give up 254,167 pieces of A, can we meet our sales requirements for that product."

Linear Programming will not make a decision for you in this situation, but it will certainly place in your hands a means of evaluating the consequences of your decision.

Forecasting The Effect of a Change in Selling Price

Now, how is Linear Programming used as a before-the-fact tool? Assume that you are progressing along the original program making none of Product B. Your sales manager forecasts that in order to stay competitive on Product A you will have to reduce your selling price to the extent that the profit per piece on Product A drops by nine cents. What should you do? Figure 5 gives the answer as developed by Linear Programming.

Product	Manufacturing Alternative			Total
	M1-M2	M1-M2A	M1-M3	Production
A	200,000	0	200,000	400,000
B	0	25,000	0	25,000
	Maximum Profit = $309,000 Sacrifice Profit = $ 41,000 Sacrifice Production = 66,667 units of A.			

Fig. 5 - Most Profitable Program if Unit Profit Drops Nine Cents on Product A.

It now becomes profitable to manufacture 25,000 units of Product B. This program yields a profit of $309,000. Your profit potential drops by $41,000 and you give up 66,667 units of Product A. Had the original program been continued (466,667 pieces of A, 0 pieces of B), the profit would have amounted to $308,000 using the lower Profit Per Piece. In this case you might decide to continue with the original program for reasons other than greatest profit since the difference between the two programs is so small. The advantage of going through the process of determining the most profitable program is that it provides the means for determining what the cost is in terms of lost opportunity. For example, further reductions in the selling price of Product A will very soon make it more profitable to manufacture Product B -- a point which may not be quickly determined otherwise.

The item of greatest importance in basing decisions upon Linear Programming is that you are always looking at your entire problem rather than part of it. Your decision is made on the basis of factual information and in terms of your over-all company profits.

A Problem Involving Distribution Costs

Not all of management's problems are as broad as the one we've just discussed. For instance, scheduling in a machine shop can be done considering only one department by itself. The following problem, illustrated by Figure 6, uses Linear Programming in a narrower sense to reduce an important item of distribution costs.

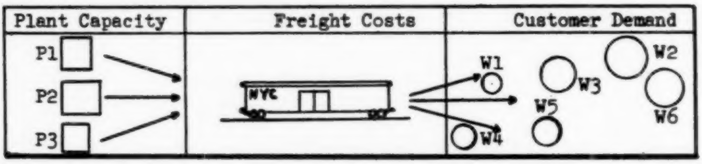


Fig. 6 - A Distribution Cost Reduction Problem

The company in question manufactures its products in three plants, P1, P2, and P3. It distributes its product through six warehouses, W1 through W6. The President of this company is faced with reducing distribution costs but he knows that today, at least, he cannot do anything that might interfere with on-time deliveries to customers. He has stated, therefore, that he wants to obtain the lowest possible freight costs and still keep his plant capacity and customer needs in balance. Figures 7 and 8 present the information we need if we are to answer this problem.

Plant	Location	Capacity per Month
1	Chicago	5,000 units
2	Boston	3,000 units
3	Atlanta	10,000 units
Warehouse	**Location**	**Monthly Customer Demand**
W1	Cincinnati	1,000 units
W2	New York	4,000 units
W3	Toronto	6,000 units
W4	Baltimore	2,000 units
W5	Knoxville	2,000 units
W6	Pittsburgh	3,000 units

Fig. 7 - Production Capacities and Customer Demands Involved in the Problem of Figure 6.

Figure 7 shows that the company's three plants are located in Chicago, Boston, and Atlanta and that their respective productive capacities are 5,000, 3,000, and 10,000 units per month.

Similarly Figure 7 presents the locations of the company's six warehouses and records the number of units per month needed by each for distribution to retailers. Cincinnati, for example, requires 1000 units per month to satisfy customer needs.

Figure 8 shows the freight costs for shipping one unit from each of the company's three plants to each of their six warehouses.

Plant	Warehouse					
	W1	W2	W3	W4	W5	W6
P1	3	3	2	5	2	1
P2	4	1	1	2	2	1
P3	2	2	5	1	1	2

Fig. 8 - Schedule of Unit Freight Costs

The cost of shipping one unit between Plant 1 in Chicago and Warehouse 3 in Toronto, for instance, is two dollars.

A Solution Based on Lowest Freight Cost

Before solving this problem by Linear Programming, let's see what would happen if we decided to make all shipments on the basis of lowest freight cost. Figure 9 presents the shipping schedule that would result.

Plant	Plant Capacity	Units Shipped To						Total Shipments	% of Cap. Used
		W1	W2	W3	W4	W5	W6		
P1	5,000						3000	3,000	60
P2	3,000		4000	6000				10,000	330
P3	10,000	1000			2000	2000		5,000	50
		Total Freight Cost = $19,000							

Fig. 9 - Shipments Based Only on Lowest Freight Cost.

This shipping schedule gives us a total freight cost of $19,000. BUT, can we meet our delivery schedule? Obviously we can't because, as the figure shows, P2 would be called upon to produce over three times its capacity.

A Logical Solution

Now let's set up another shipping schedule that stays in line with plant capacity and at the same time provides for the meeting of customer needs.

The typical way of doing this is to start with the requirements of Warehouse 1 and obtain them from the source that results in lowest freight charges. Then we would proceed to warehouses 2, 3, and so on considering both the remaining plant capacities and freight costs. Figure 10 presents the results of following this logical approach to a solution.

Plant	Plant Capacity	Units Shipped To						Total Shipments
		W1	W2	W3	W4	W5	W6	
P1	5,000		4000	1000				5,000
P2	3,000			3000				3,000
P3	10,000	1000		2000	2000	2000	3000	10,000
		Total Freight Cost = $39,000						

Fig. 10 - A Solution Arrived at by Logic -- The Typical Procedure.

One question of major importance remains unanswered, however. Can we set up a better solution and thus reduce our freight costs below the $39,000 shown in Figure 10?

The Linear Programming Best Solution

With Linear Programming, you can readily determine whether or not you can do better. And more important, you can determine a definite and specific shipping schedule which you will know is the best you can do.

Figure 11 shows this best shipping schedule that results from applying Linear Programming.

Plant	Plant Capacity	Units Shipped To						Total Shipments
		W1	W2	W3	W4	W5	W6	
P1	5,000			3000			2000	5,000
P2	3,000			3000				3,000
P3	10,000	1000	4000		2000	2000	1000	10,000
		Total Freight Cost = $27,000						

Fig. 11 - The Best Solution -- Arrived at Through Linear Programming.

You know now that under the specified conditions and limitations you cannot lower freight costs under $27,000. In addition you know how many units should be shipped from each Plant to each Warehouse in order to realize this cost.

With Linear Programming you end up with the best solution under the circumstances and limitations which you face. You can tell when you have arrived at the best solution. If you decide to follow something other than the best plan, you can measure the cost of your decision. Furthermore, your answer is, again, in terms of a definite and specific course of action. In this problem, for instance, you know exactly how many units to ship from each plant to each warehouse and you know that the best you can do in freight costs is $27,000.

One of the most interesting and useful applications of Linear Programming is a reverse application of this shipping schedule example. By doing this in reverse, it is a relatively simple problem to determine the best location of distribution or warehousing points for a given product or combination of products.

These two examples are hardly more than a hint of the wide range of management problems upon which Linear Programming can be helpful. There is, of course, no lessening of the need for good management judgment, but that judgment can now be supported with vastly expanded insight into the complications surrounding each decision. Managers can look forward to much less "guesstimate" and "seat-of-the-pants operation" in their

work, and much more fact and precision in their approach to the difficult problems ahead. We are, indeed, coming closer each day to the time when management will be less of an art and more of a science.

A STATISTICAL TECHNIQUE FOR ADJUSTING PRODUCTION TO SALES TRENDS

E. H. Robinson

Johnson & Johnson, Chicago

During the past three years we have been working in a fascinating new field of statistical application. This work involves the use of statistical techniques for making adjustments in production planning to correlate with changes in sales trends.

This new technique utilizes control limits for making decisions, and it has been applied highly successfully in our Chicago plant operations. The original technique was developed by Mrs. Frances Newman of General Electric's Electronics Division at Schenectady, New York, and we have worked with her to expand our applications. More recently, by eliminating the factor of seasonal variation, we have been able to narrow control limits so that the technique is much more sensitive and therefore, much more useful.

This presentation runs about 45 minutes and is accompanied by a group of 25 color slides which, in sequence, demonstrate the use of these methods.

APPLICATIONS OF STATISTICAL METHODS IN EVALUATING PERFORMANCE OF ELECTRONIC EQUIPMENT

Ralph L. Madison

Aeronautical Radio, Inc.

I. Background of the Aeronautical Radio, Inc. (ARINC) Electronic Reliability Program

The title of this paper may well imply a comprehensiveness which is beyond the scope of the paper itself. While I should like to discuss the application of statistical methods as a whole to the evaluation of electronic equipment performance, the fact is that many of the methods in general use in other fields have not evolved sufficiently to be of practical use in such evaluation.

Many of the statistical methods we at Aeronautical Radio -- or "ARINC," as we refer to ourselves -- have been applying are the common ones. They may be found in the standard statistical texts. However, we also face problems which do not lend themselves to solution through use of these common methods -- problems which require the modification of such methods and even the development of entirely new statistical approaches.

In this paper, I will consider some of these problems in detail. However, I should like to begin with a brief discussion of the background which provides the context within which we at ARINC have been seeking to apply statistical methodology. The work ARINC is currently doing in the electronic reliability field had its origin in an investigation of electronic tube reliability conducted for the airlines following World War II. ARINC's success in obtaining improvement in tube reliability for the domestic airlines attracted the attention of the Military, which -- as the result of several surveys conducted under its auspices -- had concluded that tubes were the major contributor to the unreliability of military equipment.

The Military engaged ARINC to conduct a field surveillance program at various military installations, the specific objectives of that program being to observe tube removals, determine the causes of such removals, and suggest means of coping with these causes. Selection of the "field surveillance" technique was based on the premise that only in practical application is there the assurance that all environmental factors affecting product life are brought into play in proper proportion.

At this point, I shall digress to make clear what is meant by "field surveillance." Field surveillance may best be understood by contrasting it with a laboratory experiment.

In the field, we may not be aware of all the environmental factors operating or of the level at which they operate. We can observe the failures resulting from this environment and try to reason back to their causes. In the laboratory, on the other hand, the environment is controlled at a known level; the product tested in this environment; and the resultant change in the product observed. Thus, in field

surveillance, we observe the result and attempt to determine the cause; in laboratory experiment, we control the cause and observe the result.

It should be noted, however, that field experimentation is possible. In evaluating electronic products, it is possible to control some environmental factors -- maintenance and operating practices, for instance -- in order to determine the effect of these factors on product life. Nevertheless, it is often quite difficult to control environment as precisely as one might wish.

At ARINC, we start with observations made in the field -- that is, at the various bases at which we are conducting surveillance programs. Our investigation thus progresses in the following chronological stages: (1) Observation of field phenomena, organization of data, and the establishment of patterns; (2) Explanation of the observed phenomena; and (3) Return to the field for verification of the explanation through planned experimentation.

These are, of course, three steps of the conventional scientific method. The statistician is concerned most directly with the first and last of these steps, inasmuch as they draw most heavily on statistical methodology. The first step -- observation of phenomena in the field -- requires that the statistical pattern or distribution describing the phenomena be adequately described. The statistician must therefore devise certain "measures" that will permit the summarization of field observations. These "measures," which must allow for ready detection of such changes that occur in the product or products under observation during the period of observation, should be easily interpretable and free of serious bias.

Simple examples of such "measures" include mean time to tube removal, mean time between equipment openings for purposes of repair, and the ratio of time spent in repairing an equipment to time during which the equipment was in trouble-free operation. There are, of course, others. Generally speaking, the problem under attack will suggest the "measure" or "measures" needed. Whether or not these "measures" can be employed to develop a probability distribution as to a given event or observation is of great importance to the statistician. If they can so be employed, the statistician can determine whether a major difference revealed in the measurement of two events may be attributed to pure chance, or -- what is more likely -- to a specific external cause.

The last of the three steps taken during the typical ARINC investigation -- the return to the field for verification of the assumed explanation -- involves the testing of our hypothesis by employment of a carefully-designed experiment. At this final stage, it is of the utmost importance to determine how large a sample must be chosen for the experiment so that any practical differences that exist in fact in the populations being sampled are highly likely to be revealed as statistically significant by the test.

Before moving on to some of the specific statistical problems we at ARINC face today, I should like to return briefly to consideration of the origins of our existing surveillance program. As I stated earlier, this program was begun after the Military had concluded that the vacuum tube was the major contributor to unreliability in electronic

equipments. On the basis of this interpretation, ARINC was charged with collecting tube removal data in the field, analyzing these data, and pointing out weaknesses in tubes to the manufacturers who produced them. It was felt that equipment reliability would be greatly improved if the manufacturers corrected these weaknesses.

I do not question the assumption that improved vacuum tubes will make for more reliable equipments, but I do question -- having had the benefit of hindsight -- that tube removals or tube removal rates are adequate measures of equipment reliability, or that optimum gains in reliability will be made solely by improvement of one type of component.

The assumption that the vacuum tube was the component in electronic equipment most often removed was correct. This conclusion has been frequently verified. The assumption that the tube removal rate is directly and immediately related to equipment reliability (or the lack of it) has not been borne out by subsequent investigation. In short, factors other than tube performance often determine the removal of tubes and thus make removal rate an inconclusive and often erroneous criterion of the reliability of the equipment in which they are employed.

It has been found, for example, that many tubes are removed because of their ease of removal and their relatively low cost, rather than because they are actually malfunctioning in the equipment. That is, the Military technician confronted with an equipment which is not operating properly will replace a tube on the premise that such replacement <u>might</u> improve the equipment's performance even though the actual trouble lies elsewhere in the equipment. Closely allied to this practice is that of removing a given tube from an equipment because a tube tester indicates that it is "weak" even though the equipment operates satisfactorily with that tube installed.

The limitation inherent in making a component study when the overall goal is equipment reliability might best be illustrated by example. Let us suppose that a particular tube type is quite frequently removed because of a cracked glass envelope. An observer who goes no further than the defective component itself would undoubtedly recommend that the tube manufacturer use stronger glass in the production of this tube type. A complete evaluation, on the other hand, might find that the tube type operated quite satisfactorily in the equipment, but that its location was such that there was a high probability of its envelope being cracked upon insertion or removal. The proper recommendation would probably call for changing the location of this particular tube type.

Our program started as a tube reliability study, but we have found it essential to continually broaden the scope of that study. Our observations and experimentation today are directed toward the determination and evaluation of the factors affecting tube life and equipment reliability.

II. <u>Specific Statistical Problems Encountered</u>

In our work, we face three basic statistical problems: (1) Definition of "reliability," so that it may be described in a quantitative manner; (2) Summarization and description of the patterns (distributions) of tube removals; and (3) Establishment of the

relationship between component (tube) removal and equipment failure.

A. Definition and Interpretation of Reliability

ARINC has proposed the following definition of reliability as it concerns an electronic product:

> "The reliability of an electronic product is the probability that the product will give satisfactory performance for a given period of time when used in the manner and for the purpose intended."

This definition, we believe, encompasses cases pertinent to our field of study. Many concepts implicit in the definition are well worth further comment.

Reliability — as we see it — implies that the equipment or component performs its function or functions satisfactorily. Inasmuch as some equipments are designed to perform more than one function, or — in the course of operation — are found to be capable of performing functions beyond that for which they were designed, we must decide whether reliability is to imply that all functions are performed satisfactorily or merely that one function is performed satisfactorily.

We must also establish what is meant by satisfactory performance of the function or functions. There are at least three criteria of satisfactory performance which may be applied in the electronic surveillance field: (1) Operator satisfaction; (2) Repair technician satisfaction; and (3) Satisfactory performance in the sense that the equipment or component meets a given specification. Inasmuch as each of these criteria will lead to a different estimate of reliability, it is important to specify which is being employed when speaking of the reliability of a given component.

Another factor which must be precisely defined in our interpretation of the reliability of a given equipment or component is that of time — i.e., the <u>period of time</u> of satisfactory performance. What constitutes a satisfactory period of operation for one type of equipment may not be for another. Further, two equipments of the same type under observation for the same period of time may differ considerably as to the amount of actual operating time accumulated.

A guided missile, for example, operates for an extremely short period of time as compared to a radar set. And the actual hours of operation may vary considerably from radar set to radar set. A ship's search radar may be required to work continuously for many days, whereas an aircraft radar may be in continuous operation for only several hours.

In determining the actual time of successful operation for an equipment, it is often necessary to take into account the number of times a given equipment or component is turned on and off — i.e., the number of duty cycles during the period of observation. It is possible that the number of on-off cycles will be considerable and thus will make for a shorter period of actual operation than might seem to be the case under casual scrutiny.

Still another factor to be considered in defining and interpreting the reliability of a given equipment or component is the amount of operating time accumulated prior to the period of observation -- if any. Generally, estimated reliability for a new equipment of a given type would differ from that for the same type of equipment with no prior operating time.

The last -- and possibly the most important -- factor to be considered in interpreting reliability is the external environment in which the product concerned is used. This environment must be defined in as precise and detailed a manner as possible.

In summary, then, the factors which must be considered in evaluating equipment or component reliability include the following:

(1) The function or functions of the equipment being considered in the evaluation;
(2) The criterion for satisfactory performance. Is it
  (a) operator satisfaction?
  (b) repair technician satisfaction? or
  (c) conformance to a prescribed performance specification?
(3) The period of time being considered as the length of trial and the number of on-off cycles occurring during this period of time;
(4) The age of the equipment at the beginning of the evaluation;
(5) The nature of the external environment in which the equipment or component is employed.

An understanding and knowledge of all of these factors is essential if the reliability of the equipment or component concerned is to be properly described.

B. <u>The Statistical Distribution of Time to Tube Removal</u>

In the electronic reliability field -- as in other fields in which statistical tools are employed -- effort is made to arrive at the parameters or characteristics of a given population from data revealed by a random sample of that population. By analysis of observed data obtained from a random sample of a given tube type in a particular environment, we may estimate the time-to-removal distribution -- or probability density function -- for the entire population of this tube type in the environment concerned.

Having arrived at the time-to-removal distribution for the population, we can estimate (1) the mean time to removal for all tubes in the population; and (2) the probability that a tube in this population will not be removed in a given number of hours -- i.e., the reliability of the tube in question. Further, we can make a quantitative comparison between the tube type under study and improved versions employed in the same environment. We can also make such comparisons between this tube type and others of the same type employed in different environments.

I should like to cite a recent experiment to illustrate the nature of the problem we face in seeking to ascertain the form of a

time-to-removal distribution for a specific tube type. We selected a random sample of 122 removals of a tube type used in a transceiver which had been in almost continuous operation. These tubes were separated into two groups on the basis of "reason for removal." We plotted mortality curves for each group and found that they differed. The tubes which had been removed because of degradation of electrical characteristics seemed to fit a mortality curve based upon the normal or possibly the gamma distribution of times to removal. Those removed as catastrophic failures or in which no defect could be found following removal, showed a mortality curve which was apparently based upon an exponential or Weibull distribution. Of the total of 122 removals, 78% were in the degradation category whereas the remainder were either true catastrophic failures or tubes in which no defect was found following removal.

This result suggests that a single time-to-removal distribution may not be adequate to describe the removal pattern for a given tube type over a long period of time. In such a case, the description might better be made by the weighted sum of several time-to-removal distributions. Such a "weighted" description might be expressed in the following form:

$$f(x) = k_1 g_1(t) + k_{2g2}(t)$$

where $k_1 + k_2 = 1$ and $g_1(t)$ might be an exponential distribution of the form $1/\Theta^{e^{-t/\Theta}}$ and $g_2(t)$ might be a normal distribution of the form

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-\mu)^2}{2\sigma^2}}$$

In this model, $f(t)$ is the probability that a tube will be removed for any reason at time t; k_1 is the proportion of all tubes which will either be catastrophic failures or for which no defect will be found following removal; $g_1(t)$ is the probability that a tube suffering a catastrophic failure or one in which no defect is found will be removed at time t; k_2 is the proportion of all tubes which will be removed as degradation failures; and $g_2(t)$ is the probability a tube which is a degradation failure will be removed at time t.

Given a time-to-removal distribution of this type, we are in a position to estimate its parameters from a sample of the <u>observed</u> times to removal. Inasmuch as the distribution is actually the sums of two distributions — one exponential and the other normal — we would like to make estimates of k_1, k_2, Θ, μ, and σ. If our sample of observed times to removal were truly random, these parameters could be estimated by familiar methods.

However, it has been our experience that very few samples of observed times to removal are random — random in the sense that a tube removal at 10,000 hours of operation has the same chance of appearing in the sample as one at 50 hours of operation. Generally speaking, our ARINC observations extend over a fixed calendar time, e.g., one year. The tubes under observation are installed in several different equipments, each of which may be operated a different length of time during the period of observation. Thus, tube removals occurring after the end of the observation period are not accounted for in the sample. Further, the operating time accumulated during that period having varied from equipment to equipment, our sample of times to removal is a truncated one, or — more generally — a multiple truncated one.

In such a multiple truncated sample, we have complete time-to-removal information on only the removals occurring in the period of actual observation. As for those tubes not removed during the period of observation, we know only their time of actual operating during the period. Therefore, if we are to estimate mean time to removal for any tube type in the equipments under observation, we must make an assumption about the manner in which tubes not removed during the period of observation would have been removed had that period been extended.

If we assume that the removals of all tubes in the sample will fit an exponential probability distribution, the estimate of Θ, the mean time to removal for the sample, is expressed as:

$$\hat{\theta} = \frac{\sum_{i}^{r} t_i + \sum_{j}^{m} N_j T_j}{r}$$

where t_i is the time to the i^{th} removal and n_j the number of tubes still in operation at time t_j.

There are statistical methods available for estimating the parameters μ and σ for a normal distribution of times to removal given a sample which has only one truncation time -- i.e., all tubes not removed during the period of observation are known to have operated the same length of time. However, we have found no completely satisfactory method for estimating these parameters on the basis of a sample with several times of truncation. Nor have we -- given such a truncated sample -- devised a method of estimating k_1, the proportion of all tubes installed which would eventually be removed as catastrophic failures or for which no defect would be found following removal, and k_2, the proportion of all tubes installed which would eventually be removed as degradation failures.

When dealing with a multiple truncation sample of time-to-removal, the usual procedure has been to assume that all the tube removals involved would fit the exponential distribution pattern. This assumption makes possible the simple estimate of mean time to removal, Θ, previously cited. However, we use this approach with great caution.

The problem which may present itself when basing an estimate upon this assumption is typified by the results achieved when the assumption is applied to the sample of 122 tube removals noted earlier in this discussion. The true mean time to removal for this sample is 3,500 hours. Had we assumed that all of the removals in the sample fitted the exponential distribution and had we observed only those removals made during the first 2,000 hours of operation, our estimate of the mean time to removal for the sample would have been 5,750 hours. As you can see, this estimate would have been 1.6 times greater than the true value.

It is clear, I think that the example of a multiple truncated sample which I have just cited is not an isolated one. We must expect to encounter multiple truncated samples of removals of other tube types in other applications in which the degradation removals fit one type of distribution and the catastrophic removals another. And we must continue to be alert to the possibility that a serious error can be made in estimating the mean time to removal for a multiple truncated sample by assuming all failures to fit the same exponential distribution.

What is needed are: (1) More detailed study of tube removals to determine the kinds of probability distributions which describe the various types of removals (degradation, catastrophic, etc.); and (2) Methods of estimating the parameters of these distributions from multiple truncated samples.

C. <u>The Relationship Between Component Removals and Equipment Failure</u>

Assuming that we were able to determine the exact time to removal distribution for components, we should like to use this information to determine the time to failure distribution of equipments using these components. It might thus be possible -- at the drawing-board stage of equipment production -- to estimate the reliability of the proposed equipment and also to determine what components and quantities should be used to maximize its reliability.

Whenever we have tried to determine equipment reliability from our knowledge of components time to removal distributions, our estimates have been in error when checked against observed equipment reliability. Generally, an estimate of equipment reliability based upon component time to removal distribution will be less than the true equipment reliability. Our difficulty seems to stem from two sources: (1) The interdependency of the components within the equipment; and (2) The validity of the time to removal distribution for the component.

Let us assume that we have two equipments, each performing the same function in the same environment and using the same components in the same quantities. Let us further assume that the designs of the equipments differ. We would not expect the reliability of these two equipments to be exactly the same because the reliability of an equipment is not only dependent upon component reliability, but also upon equipment design.

Component time-to-removal distributions are also affected by component interdependency. This dependence of the component time to removal distribution on equipment design is evident in the difference between the removal rates of two sockets within the same equipment using the same tube types. If equipment design did <u>not</u> affect component time to removal, then the removal rates from both sockets should be similar -- which is seldom the case. In order to more precisely estimate equipment reliability from component time to failure distributions, we must devise methods of incorporating the design factor into our methods of estimation.

We have sought to compensate for equipment design as a factor influencing the time-to-removal distribution of a component by attempting to estimate the reliability of the type of equipment in which the <u>particular</u> component is employed. However, using this procedure -- i.e., estimating the reliability of a given type of equipment from the time-to-removal distributions of components removed from that type of equipment -- leads to a conservative estimate of equipment reliability.

The reason for this conservative estimate or underestimate seems to be two-fold: First, not all removals are those of failed components. Some of the components removed are good in the sense that they will work in the equipment when reinstalled. Second, not all components removed as failures have actually caused equipment failures.

A component may fail, cause equipment failure and -- at the same time -- cause other components to fail.

While the statistical technique for estimating equipment reliability from the time-to-removal distributions of the components seems mathematically sound, we must account for equipment design and environment as factors affecting equipment reliability. An environment and application study aimed at this objective is under way.

III. Summary

In this paper, I have tried to outline ARINC's approach to the study of reliability and to sketch some of the background developments which have made the study possible and which have contributed to the progress we have made to date. And we have made progress.

But, we also admit to having made errors and to being confronted by a number of problems, a few of which I have discussed here.

Specifically, I have dealt with three basic problems of a statistical nature:

A. The detailed definition of "reliability" so that it may be described in a quantitative manner;

B. Accurate determination of the time-to-removal distributions of electron tubes; and

C. Determination of the specific relationship between the removals -- that is, of time-to-removal distributions for such removals -- and equipment "failure."

We feel that we have been and are moving in the right direction toward satisfactory solution of each of these problems. We also realize that we still have a long way to go.

CONTROL CHARTS IN MULTI-STAGE BATCH PROCESSES

R. S. Bingham, Jr.
Atlas Powder Company

Application of statistical control in the chemical process industry has been hampered partly by lack of understanding of how statistical concepts may be applied to advantage beyond that commonly gained with standard chemical analyses and physical tests. Renner (17) has pointed out reasons for reluctance among chemists to adopt statistical control tenets, while Bicking (2,3,4), Wernimont (20), and Hader & Youden (12) have discussed scores of problems solvable with the techniques. This paper reviews underlying concepts of statistical control as applied to batch chemical processes. A study of spent acid reduction in a counter-current 3-stage nitration process illustrates some of the principles.

Interpretation of Concepts for Batch Control

In the chemical industry or any other field, no control is valid unless it means regulation within important, technically selected limits. In most cases, the limits are pre-set from background knowledge to meet customer requirements, or from economic restrictions; but process capability may be the governing factor. The limits must be set to reflect deviations associated with departures from standard operating procedures of practical importance, and not to call attention to fluctuations in raw materials, operating conditions, or operator manipulations within the allowable range. Limit exceedances call for corrective action - adjustment of the process, recycle or reworking of the material under reaction. For batch processes, other prerequisites of the "statistical control" concept such as observation order, randomness and rational subgrouping, are incorporated in the sampling and charting instructions peculiar to each situation. Justification for chart control includes recognition that the technique provides psychological as well as technical advantages for guiding operations, whether these processes include raw material acceptance, manufacturing, laboratory analysis, product verification, or experimentation. Although the chemical phase of processing control is generally the main consideration, attention to physical entities like time cycles, ingredient weights, and equipment calibration reduces chance variation, further increasing detection sensitivity.

Types of Batch Control

The type of batch control used depends on technical knowledge of the process and available tests. Charges may be considered individually or treated as parts of a larger semi-continuous process. "Within-batch" control may utilize charts to follow a reaction to an end point. A regression line and limits generally replace the conventional central line (14). Figure 1 shows a typical case. Specifically, individual results taken over the time period of interest are plotted against the given line and its 2-standard deviation limits. For good power, the test method must be quick compared to the batch cycle and possess high reproducibility. If the batch is to be "cooked" to reach a given end point, thus influencing cycle time, the method is limited to single-vessel reactions and co-current flow sequences.

"Batch-to-batch" control (13) treats successive charges as parts of a continuous procedure in which restraints on operator methods, chemical

variables, and producing conditions assure reproducibility. Although making good batches remains the goal, the guiding philosophy makes use of information from preceding batches for control of processing conditions for succeeding charges. Operator aids in the form of instrumentation and chart control must be provided to avoid assignable causes from the human element. As in "within-batch" control, single samples from well-mixed vessels of liquids or gases may be compared on control charts for individuals. Choice of probability factors depends on power needed; but 2-sigma limits for individuals and moving range charts, combined with run theory, seem adequate for many situations (19). If test methods are too variable, replicate tests on the same sample may be needed for adequate precision (1). Solids or aggregates of various sizes may require a detailed sampling plan to give representative results (15). The control chart for individual batch results follows changes in the process average for batches made under the same conditions. The moving range chart detects increases or decreases in batch-to-batch uniformity, trends, and serial correlation (see Figure 1). The relative insensitivity of the individual batch chart compared with an average chart for detecting shifts in average can be overcome by proper choice of control limits and sampling frequency, often limited to one sample per batch. Moving averages over several batches may be used where the analogy to physical compositing seems appropriate. Generally, converging parallel process streams with subsequent common treatment or counter-current stepwise flow practices are suited to "batch-to-batch" type control.

A Decision For Chart Control

Chemists and chemical engineers have provided process control without control charts by using instrumentation, operator log sheets, chemical analyses, and physical tests. To justify chart control, convincing arguments must be presented citing advantages of graphic presentation, action limits for operators, rapid detection of improvements or losses, awareness of trends or changes in variability, and ease of correlation of processing variables. The continual comparison of "what the process is doing" with its capability or established standards is apt to be most rewarding. The achievement of greater product uniformity by the elimination of unnecessary processing adjustments, commonly made by shift crews when taking over the shift, helps to pay the cost of control. Tighter control of trace impurities and complex chemical reactions will strengthen the case for chart control (7,9,11).

Control of Multi-Stage Processes

Multi-stage processes frequently have an additional level of complexity from the viewpoint of control. Usually, the process involves raw material additions at several stages, changes of state, and reaction in a variety of equipment; all of which may influence control. In co-current flow, control at early stages allows corrective action in subsequent steps providing processing conditions are favorable. A continuous check on product quality is possible by following a particular batch from stage to stage observing various check-point results.

In counter-current flow, process information must travel both directions for control. In many cases, where isomer formation or yield is influenced by operating conditions at each of the various steps, control of reactants must be accomplished at each stage even though over-all chemical "balance" is maintained. Considering the method of operation it is

not surprising to find that control of a particular batch in the later stages of the process eventually produces control of a batch six or seven steps behind. These characteristics emphasize the continuous nature of the process over that of the individual batches. It should be realized that in many cases the continuous process corresponding to the multi-step batch process is a limit to be approached more closely as control improves. The physical analogy to this comparison is apparent since many batch processes would be continuous processes if economics, equipment design, available equipment, or reaction kinetics permitted.

An Example

The principles described may be illustrated with some data taken on a three-stage batch nitration process using counter-current flow of nitrating acid and organic (8,10,18). The material flow is shown in Figure 2. Acid concentrations, temperatures, and reaction times are necessarily increased as the nitration adds additional "nitro" groups to the organic molecule. After strengthening, waste acid from the last stage is used in the second stage and waste acid from the second stage, in the first stage. Routine control of the process covers product quality, yield, reactant use, reactant recovery, and acid strength.

From a review of production records, the operating department decided that both the amount of nitrating acid charged per batch and the amount to be recovered could be reduced with a resultant saving in reactant and possible increase in plant capacity. Calculations were made by the technical department to determine the desired operating point. The amount of nitrating acid to strengthen the waste from the second stage was reduced for one production line during a ten-day plant trial. After successful results for two days, the same procedure was instituted on other production lines. Several days later, the operating department reported that the yield was dropping. The test was halted and a data analysis requested to evaluate, if possible, whether the desired reduction had been achieved, whether the test was sufficient to preclude the reduction as a realistic possibility again, how much the yield had been reduced, and any other relations among the variables measured.

Since the plant test had been carried out without control charts as guides, the daily averages for each of the measured variables were plotted as individuals. Daily averages were used for comparison since yield data on a batch-by-batch basis was not considered reliable. Control limits for the individual batch and moving range charts were calculated from the average and standard deviation for each of the two time periods - "before" and "during the test." (These 2-sigma limits were shown on the charts even though insufficient points were present to establish limits; and lack of control was noted in the form of trends, runs, and single points out of control.) Confidence limits for the averages and variances were calculated using the true degrees of freedom to judge shifts in average or changes in dispersion.

Prior to data evaluation a review of reaction chemistry was made to establish tentative hypotheses of interest that might be tested. It was postulated that even though an acid reduction at the first stage might reduce the degree of reaction, the second stage might be adequately acid-rich to complete both reactions. In some cases, agreement could not be reached as to which were the controlling variables. To resolve the problem, a multiple regression was calculated to determine the influence of

"mono oil" weight (x_1), "bi oil" weight (x_2), amount of fortifying acid for mono-nitrating acid (x_3), and amount of fortifying acid for bi-nitrating acid (x_4) on yield (y). Since the chemistry might change in the first stage after the acid reduction, correlations were tested in two parts - using data taken during the control period before the plant test, and data covering both the control and test periods.

From the control charts, shown in Figures 3 and 4, and the multiple regression and correlations summarized in Table I, it was concluded that:

1. No statistically significant change in yield occurred.

2. The weight of bi waste fortifying acid had been reduced and was controlled at the desired level, with the exception of the last 2 days.

3. The acid concentration in the mono waste was materially reduced and controlled at a lower level. A least squares equation was derived for predicting mono waste concentration from weight of bi waste fortifying acid that agreed reasonably well with theory.

4. Mono oil weight and bi oil weight were significantly related during the pre-test period; however, the relation was not significant for the data covering both the pre-test and test periods. The significant drop in mono oil weight related to the reduction in bi waste fortifying acid apparently did not interfere with second stage reaction and over-all yield.

Furthermore, it was apparent that sufficient information had been developed from the data analysis to guide another plant test. From the relations identified, adequate predictions could be made to quickly detect whether any future test would perform according to schedule. Control charts for data evaluation sold themselves both as experimental and process control tools.

In retrospect, control charts contribute to batch process control and plant experimentation by:

1. Graphically portraying ranges within which variation is to be expected, and hence providing limits for action, thus eliminating misunderstanding about what constitutes significant modifications (16).

2. Focusing attention on natural process variation present in the process prior to change.

3. Questioning whether data collected are suitable for control. (In the above example, calculations based on cumulative yields were discarded since the variance of daily reported yields decreases from the first of the month to the last. On this basis, the date of a plant test would partially determine its outcome.)

In the particular example described above, control charts were not used for "within-batch" process control since no sufficiently rapid chemical analysis was available. Instead, "batch-to-batch" control was used

to guarantee product quality.

Summary

Each type of batch control has a part to play in modern multi-stage control. Prior to selection, a balance must be made between cost of information gained and its control value (5). More often than not the variables to be controlled for optimum regulation are not apparent. Multiple regression techniques and experimental designs (6) may be necessary to identify controlling factors. After identification, control charts may be used for daily checks on operating levels and as operator guides. In simplest form, control charts have justified their use if they awaken chemists and engineers to the natural variation of processes and lead to better control.

FIGURE I- LEVEL, UNIFORMITY AND END POINT OF BATCHES

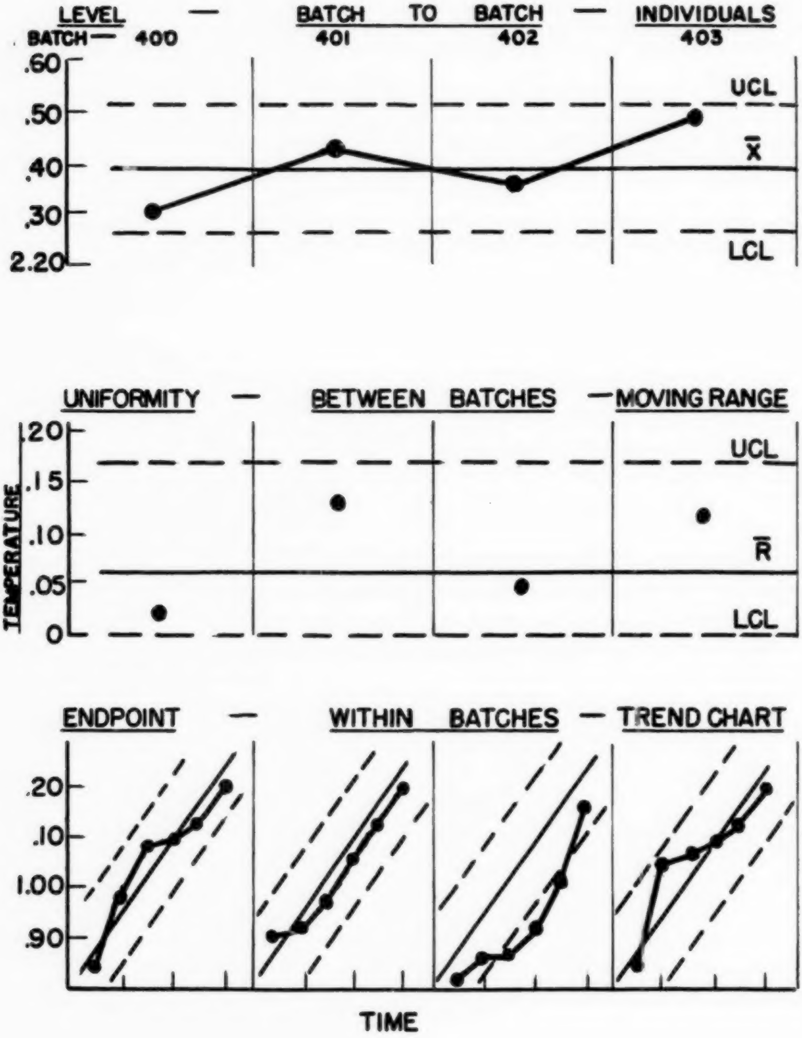
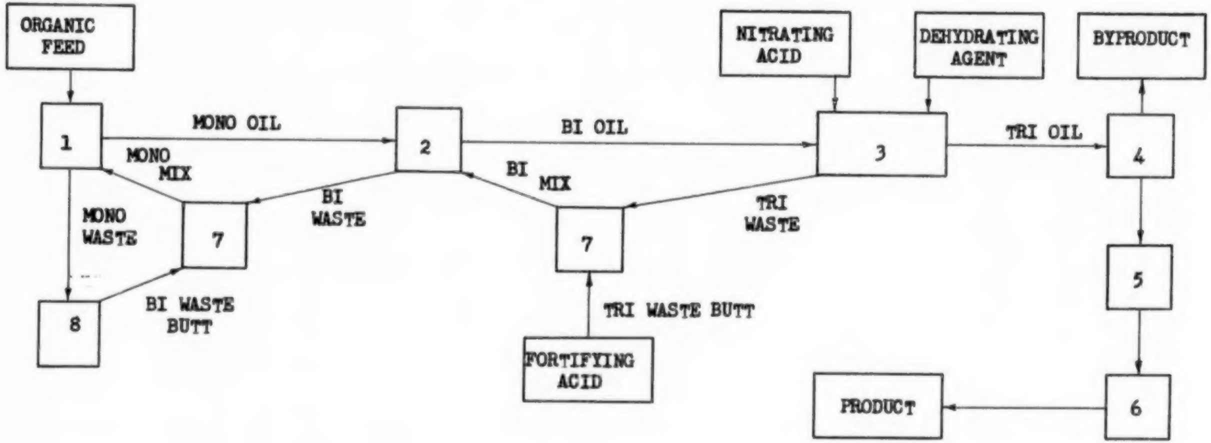


FIGURE 2 - PROCESS FLOW SHEET



KEY: 1. MONO NITRATION
2. BI NITRATION
3. TRI NITRATION
4. PURIFICATION
5. DRYING
6. PACKAGING
7. FORTIFYING WASTE ACIDS
8. RECOVERY

TABLE I - MULTIPLE REGRESSION AND CONTROL CHART EVALUATION

Multiple Regression

Variables		Data		Percent of Variance Explained	Equation Slope
Dependent	Independent	Before Test#	All		
y	x_3	*		47	- .244
x_2	x_1	**		90	.948
x_1	x_3		**	39	.142
x_5	x_3		***	81	.00449

Control Charts

Chart	Change in Average	Comments
x_1	*	Trend before plant test
x_2		Trend before plant test
x_3	*	
x_4		
x_5	**	
y		

* Significant at .05 probability level.

** Significant at .01 probability level.

*** Significant at .001 probability level.

\# See Table II

TABLE II - MULTIPLE REGRESSION SOLUTION - "BEFORE TEST" DATA

Corrected Sums of Squares and Cross Products

$Syx_1 = 645.246$	$Sx_1^2 = 4834.1$	$Sx_1x_2 = 4582.60$	$Sx_2x_3 = 2846.4$
$yx_2 = 479.596$	$x_2^2 = 7113.6$	$x_1x_3 = -\ 803.1$	$x_2x_4 = -\ 812.0$
$yx_3 = -\ 2169.376$	$x_3^2 = 12194.1$	$x_1x_4 = -\ 1047.$	$x_3x_4 = 1027.0$
$yx_4 = -\ 422.120$	$x_4 = 890.0$	$Sy^2 = 818.7404$	

Original Matrix

b_1	b_2	b_3	b_4	c
4834.1	4582.6	- 803.1	- 1047	645.296
	7113.6	2846.4	- 812	479.596
		12194.1	1027	- 2168.376
			890	- 422.120

Auxiliary Matrix

4834.1	.947974	- .166132	- .216586	.133437
4582.6	2769.42	1.30288	.0651953	- .0476309
- 803.1	3607.72	7360.27	.083944	- .256835
- 1047.0	180.528	617.853	736.767	- .156259

Final Matrix

$b_{1.} = -\ .2064$
$b_{2.} = .2801$
$b_{3.} = -\ .2437$
$b_{4.} = -\ .1563$

$B_1 = -\ .502$ (3)
$B_2 = .826$ (2)
$B_3 = -\ .940$ (1)
$B_4 = -\ .169$ (4)

Variance Analysis

Source	D. F.	S. S.	M. S.	F
Explained by $b_{3.}$	1	385.94	385.94	8.66 *
Explained by adding $b_{1.}$, $b_{2.}$, and $b_{4.}$	3	209.87	69.96 #	1.57
Residual	5	222.930	44.59	
Total	9	818.740		

Even if the entire S. S. were assignable to one variable alone, it would not be significant.

FIGURE 3 - CONTROL CHARTS

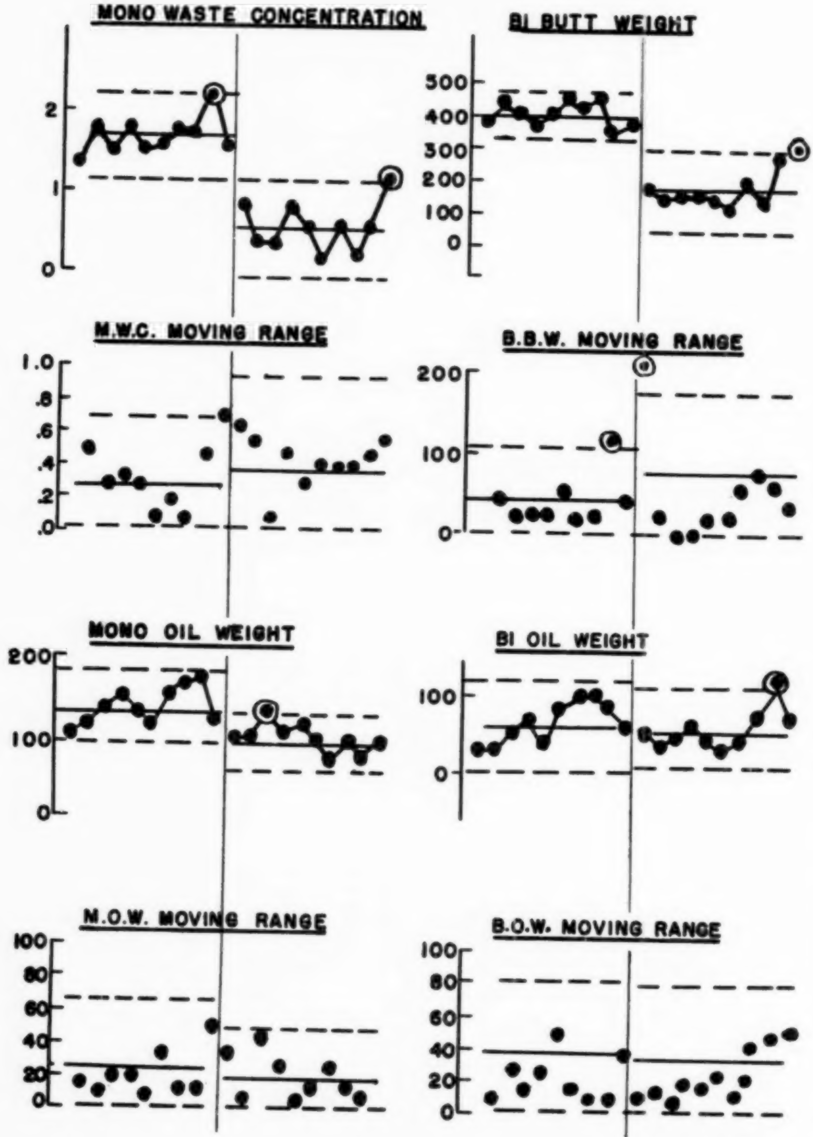
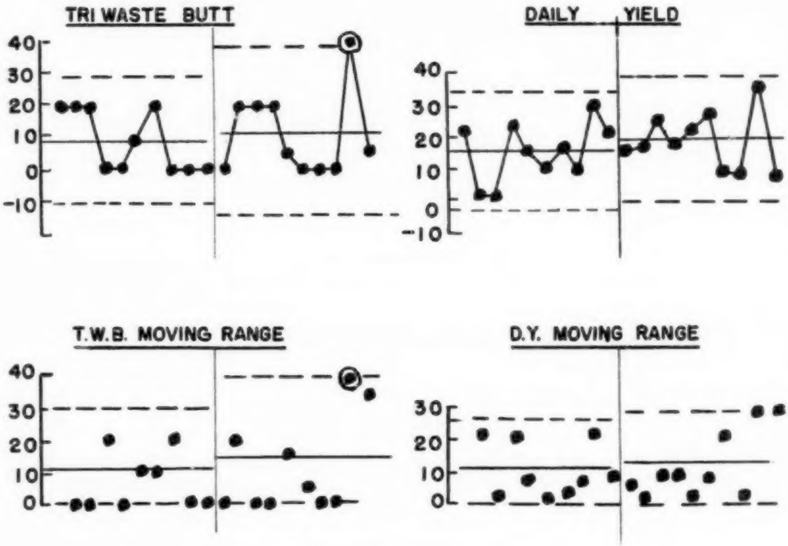


FIGURE 4-- CONTROL CHARTS



REFERENCES

(1) Bennett, C. A., "Effect of Measurement Error on Chemical Process Control", Industrial Quality Control X (4) 1-54, pp 17-20.

(2) Bicking, C. A., "Quality Control in the Chemical Process Industries" in Quality Control Handbook, J. M. Juran, Editor, McGraw-Hill Book Co., Inc., 1951, pp 561-612.

(3) Bicking, C. A., "How Statistics Contributes to Quality Control in the Chemical Industry", Industrial Quality Control, 1-47, pp 17-20.

(4) Bicking, C. A., "Statistical Techniques for Production Control and Plant Scale Experimentation", Industrial Quality Control, 5-47, pp 11-14.

(5) Bingham, R. S., "Practical Process Control", Quality Control Clinic Papers, 1954, pp 47-56, Rochester Society for Quality Control.

(6) Brownlee, K. A., Industrial Experimentation, 3rd Am. ed. Brooklyn, N. Y., Chemical Publishing Co., Inc., 1949.

(7) Ferencz, P. & Lloyd, B. H., "Statistics Can Help in Solving Process Problems", Chemical Engineering, 5-53, pp 207-210.

(8) Fieser, L. F. & Fieser, M., Organic Chemistry, D. C. Heath & Co., Boston, 1944, p 578.

(9) Gnaedinger, P. E., "Attaining Uniformity in Processing Rubber Batches", Quality Control Conference Papers, 1953, pp 55-64, A.S.Q.C.

(10) Groggins, P. H., Unit Processes in Organic Synthesis, 3rd ed., New York, McGraw-Hill Book Co., Inc., 1947, pp 1-72.

(11) Gross, M., "Q. C. in the Synthetic Rubber Industry", Industrial Quality Control, 11-47, pp 11-16.

(12) Hader, R. J. & Youden, W. J., "Experimental Statistics", Analytical Chemistry 24 (1), 1-52, pp 120-124.

(13) Hinchen, J. D., "Control Charts in Batch Processes", (New England Quality Control Conference Papers, 8-22-52?), pp 113-120.

(14) Hinchen, J. D., "Correlation Analysis in Batch Process Control" Quality Control Convention Papers, 1954, pp 101-113, A.S.Q.C.

(15) Lloyd, B. H., "The Statistical Approach to Bulk Sampling Problems", Industrial Quality Control IX (7), 5-53, pp 113-120.

(16) Mitchell, J. A., "Experiments in the Chemical Plant", Quality Control Conference Papers, 1953, pp 29-38, A.S.Q.C.

(17) Renner, M. S., "The Role of Statistical Quality Control in Today's Chemical Company", Quality Control Convention Papers, 1954, pp-515-522, A.S.Q.C.

REFERENCES (CONTINUED)

(18) Shreve, R. N., The Chemical Process Industries, 1st ed., New York, McGraw-Hill Book Co., Inc., 1945, pp 448-449.

(19) Weiler, H., "The Use of Runs to Control the Mean in Quality Control" Journal American Statistical Association 48 (264), 12-53, pp 816-825.

(20) Wernimont, G., "Statistics Applied to Analysis", Analytical Chemistry 21 (1949), p 115.

SOME ELEMENTARY THEORY OF STRATIFICATION

W. Edwards Deming
Graduate School of Business Administration
New York University

PART A. PURPOSE AND GENERAL REMARKS

The purpose of stratification. Stratification is a scheme for making use of information that is already in our possession (as from the last census), or that is obtainable at a cost not too great (as by preliminary tests or interviews) concerning some of the characteristics of some of the sampling units in the frame, the aim being to attain greater precision than would be possible for the same cost without stratification; or, alternatively, to attain the same precision for less cost.

Stratified sampling has many meanings. First of all, there are many ways to classify a sampling unit, as by source, raw material, geographic position, average rent in the area (in the case of economic surveys), size of city, density of population, proportion colored, type of predominant industry or of agriculture, etc. Second, for any given system of classification, there are several ways to draw the sample and to select the units for interview or for test, and several ways to make the estimates. One must select out of these numerous possibilities one that shows promise of being more precise than another for the same cost. Theory and experience form the only safe basis for this decision.

When one thinks of stratified sampling, he must consider and compare several main avenues of procedure:

Plan A. Don't stratify at all. This is sometimes the best plan of all.

Plan B. Classify all the sampling units of the frame. Then use proportionate allocation.

Plan C. Classify all the sampling units of the frame. Then use Neyman allocation.

Plan D. Classify only the sampling units in the sample, not the whole frame. In the formation of the estimates, force the proportions (weights) P_i to agree with known values. (P_i is the proportion of the sampling units in the frame that belong to Stratum i.)

Plan E. Classify one by one only the sampling units in a preliminary sample until you reach certain preassigned sizes of sample (n_i) from all the various strata. Discard any unit that belongs to a stratum whose quota is already filled. The sizes n_i of the samples will be fixed by proportionate allocation. Here the weights are forced in advance.

Plan F. Classify one by one only the sampling units in a preliminary sample of a designated size. Thin the samples by ratios dictated by the Neyman allocation.

Plan G. Classify only the sampling units in a preliminary sample, as in Plan E, but with the sizes (n_i) fixed by the Neyman allocation.

Any of these plans may be used to estimate the proportions of individuals in classes finer than the original strata; and these ratios may often then be used advantageously to form estimates of totals (called ratio-estimates, mentioned later).

The aim of studying the theory of stratified sampling. With the help of theory, one may make a sensible choice of plan. He will be able to dismiss from consideration those plans for stratified sampling that would show but little gain in precision, or which would raise costs considerably. A little theory will conserve funds and take the place of a vast amount of experimentation.

> Remark. Theory for the comparisons of variances will help to determine which plan is likely to be most efficient under any given set of costs. Besides cost, one must consider (1) speed, (2) possession or availability of information by which to classify a sampling unit, (3) knowledge and experience of the people who will do the work, (4) personal preference.

A simple example of a random allocation. If we define certain strata, as by geographic location, size of city, proportion colored, etc., but draw the sample at random from the entire frame, ignoring the strata, the sample-sizes that fall into the various strata will be random variables. We shall look at a simple illustration in two strata.

Zone 1 consists by definition of 70 specified squares, and Zone 2 consists of the remaining 30 squares. Let us draw a sample of 20 squares from the 100 squares (the frame), to see how they distribute themselves between the two zones. If each zone contributed its proportionate share to the total sample, then:

Zone 1 would contribute	14 squares
Zone 2 would contribute	6 squares
Total	20 squares

Now let us see how the sizes of sample distribute themselves in one particular trial. We open our table of random numbers (Kendall & Smith, 23d thousand, cols. 23 and 24, line 17, where I had stopped a few days ago on a sampling job), and read out:

39	68	69	11	32	36
17	24	96	79	95	44
09	20	12	25	92	43
37	00	19	53	31	91
29	24	90	28	11	38

These random numbers struck 16 squares in Zone 1, and 4 squares in Zone 2. So we may write, for this one trial,

$$n = 16$$
$$\underline{n = \ 4}$$
$$n = 20 \text{ (fixed)}$$

PART B. SAMPLE-SIZES FIXED IN ADVANCE: PROPORTIONATE ALLOCATION, PLAN B

Reasons why proportionate allocation to strata (Plan B) may show a gain. In the first place, stratified sampling is possible only when information exists already by which to classify any sampling unit in the frame, or when one can procure such information, inexpensively, as in data published for small areas, or by quick interviews.

We saw in the previous illustration what happened when we let the sample fall at random over the whole frame, unstratified. The random numbers fell into the two strata nearly but not quite in proportion to the number of sampling units in the two strata. In other words, the sample-sizes were not proportionate. The random failure of proportionality, and the fact that the means of the two strata may be unequal, are the basic reasons why proportionate stratification sometimes achieves better precision than no stratification at all.

Remark 1. It might seem at first thought that this failure of proportionality would cause but very little loss in precision, because (as the total sample is fixed) what one stratum loses, the other gains. However, the gain and the loss do not exactly offset each other. Thus, in the illustration, Zone 2 lost 2 sampling units to Zone 1; this loss was 1/3d of the expected sample-size (6) in Zone 2, but was a gain of only 1/7th of the expected sample-size (14) in Zone 1. If the means of the two strata are unequal, some loss in precision will result.

Remark 2. N is the total number of sampling units in the frame, for all the strata combined. n is the total number of sampling units drawn into the sample, from all the strata. $\bar{N}$ and $\bar{n}$ are averages per stratum.

Some relations between the standard deviations in the table.

$a = P_1a_1 + P_2a_2 + P_3a_3 = \frac{A}{N}$ — the average population per sampling unit (1)

$\bar{\sigma}_w = P_1\sigma_1 + P_2\sigma_2 + P_3\sigma_3$ — the weighted average standard deviation within strata (2)

$$\sigma_w^2 = P_1\sigma_1^2 + P_2\sigma_2^2 + P_3\sigma_3^2 \qquad \text{the weighted average variance within strata} \tag{3}$$

$$\sigma_b^2 = P_1(a_1 - a)^2 + P_2(a_2 - a)^2 + P_3(a_3 - a)^2$$

$$= P_1a_1^2 + P_2a_2^2 + P_3a_3^2 - a^2 \qquad \text{the variance between strata} \tag{4}$$

$$\sigma^2 = \sigma_b^2 + \sigma_w^2 \qquad \text{the total variance} \tag{5}$$

Formulation of the gains in proportionate sampling. We shall now formulate analytically the difference between no stratification and proportionate stratification (Plans A and B). In the first place, we need a mathematical definition for the proportionate stratified sampling in Plan B, which will be this:

$$n_1 : n_2 : n_3 : \bar{n} : n = N_1 : N_2 : N_3 : \bar{N} : N \quad \text{(Plan B)} \tag{6}$$

or

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{\bar{n}}{\bar{N}} = \frac{n}{N} \tag{7}$$

stated otherwise,

$$n_i = n\frac{N_i}{N} = N_i\frac{\bar{n}}{\bar{N}} \tag{8}$$

The variance of X for Plan A is

$$\text{Var } X = \sigma_X^2 = N^2(1 - \frac{n}{N})\frac{\sigma^2}{n} \tag{9}$$

We may now form the estimate

$$X = X_1 + X_2 = \frac{N_1 S_1}{n_1} + \frac{N_2 S_2}{n_2} \qquad (10)$$

and use Eq. 9 for the variances of X_1 and of X_2 in turn. For any sample in which we know N_1 and N_2, and for which we fix n_1 and n_2 in advance,

$$\text{Var } X = N_1^2 (1 - \frac{n_1}{N_1}) \frac{\sigma_1^2}{n_1} + N_2^2 (1 - \frac{n_2}{N_2}) \frac{\sigma_2^2}{n_2} \qquad (11)$$

In proportionate sampling Eq. 10 reduces to

$$X = N \frac{S}{n} \qquad \left[\text{Proportionate sampling, Plan B. } S = S_1 + S_2 \text{, the total population of the sample}\right] \qquad (12)$$

and Eq. 11 reduces to

$$\text{Var } X = N^2 (1 - \frac{n}{N}) \frac{P_1 \sigma_1^2 + P_2 \sigma_2^2}{n}$$

$$= N^2 (1 - \frac{n}{N}) \frac{\sigma_w^2}{n} \qquad \left[\text{Proportionate sampling; Plan B}\right] \qquad (13)$$

Remark 1. The reader should note that Eq. 12 is the same as the estimate X for Plan A, no stratification. The symbol S represents the population in the sample, stratified or not.

Remark 2. For this reason, a proportionate sample is a self-weighted sample, although there are other types of self-weighted samples, some of which we shall encounter. In a self-weighted sample, no weighting is required: we merely pool the results from the several strata to form S, and multiply S by N/n to form the estimate X. A computer need not be aware of the fact that the sampling was stratified.

Remark 3. For most purposes, a self-weighted sample achieves nearly the maximum efficiency. Exceptions occur, and the Neyman allocation will be an example. The reader will perceive that in Neyman sampling he can not pool the results from the strata until he has formed the estimates X_1 and X_2 separately.

Remark 4. Eq. 13 for the Var X in proportionate sampling has the same form as Eq. 9 for Plan A, unstratified, except for σ_w^2 in place of σ^2. Thus, in proportionate sampling we eliminate the effect of the differences between the means of the strata, as is obvious from the fact that $\sigma^2 = \sigma_w^2 + \sigma_b^2$.

PART C. SAMPLE-SIZES FIXED IN ADVANCE: NEYMAN ALLOCATION, PLAN C

Neyman allocation to strata (Plan C). One may be able in some problems to improve on proportionate sampling by altering n_1 and n_2 in proportion to σ_1 and σ_2. This is so when it is possible to form strata so that their variabilities (as measured by σ_1 and σ_2) are distinctly different. Such a plan was first put into practice by Neyman. Two strata will be sufficient for illustration of the theory. We start with

$$\left.\begin{aligned} n_1 &= n\frac{N_1\sigma_1}{k} + h \\ n_2 &= n\frac{N_2\sigma_2}{k} - h \end{aligned}\right\} \qquad (14)$$

The reader should satisfy himself that, no matter what be h, n_1 and n_2 when added together will give n, the total sample, provided

$$k = N_1\sigma_1 + N_2\sigma_2 = N\bar{\sigma}_w \qquad (15)$$

Later, we shall see that when h = 0 the above equations give the Neyman allocation. We wish to see what happens to the Var X for different values of h, including h = 0. When we solve this problem we shall not only discover the optimum allocation, but also how much precision we lose by making an approximate Neyman allocation (as we can only do in practice) instead of an exact one.

So now let us substitute the above values of n_1 and n_2 into Eq. 11 which is valid for any fixed allocation of the sample into the strata. Here is what we get:

$$\text{Var } X = \frac{k}{n}\{N_1\sigma_1 + N_2\sigma_2 + 0 + \frac{h^2k}{nn_1} + \frac{h^2k}{nn_2}\} - N\sigma_w^2$$

$$= \frac{k^2}{n} \{1 + \frac{h^2}{n_1 n_2} \} - N \sigma_w^2$$

$$= N^2 \frac{(\bar{\sigma}_w)^2}{n} \{1 + \frac{h}{n_1} \frac{h}{n_2} \} - N \sigma_w^2 \qquad (16)$$

Here we have a very important result.

1. The term $h^2/n_1 n_2$ is positive whether h be positive or negative; it is 0 only if h = 0.

2. Therefore, Var X is at its minimum if h = 0. Now if h = 0, Eq. 14 gives what we shall call the Neyman allocation, defined as

$$n_i = \frac{N_i \sigma_i}{k} n \qquad \text{[Neyman allocation]} \qquad (17)$$

or

$$n_i : n_j = P_i \sigma_i : P_j \sigma_j = N_i \sigma_i : N_j \sigma_j \qquad (18)$$

wherein k has the value shown in Eq. 15. Eq. 16 then shows that the minimum or Neyman variance is

$$\text{Var } X = N^2 \{ \frac{(\bar{\sigma}_w)^2}{n} - \frac{\sigma_w^2}{N} \} \qquad \text{[The Neyman variance]} \qquad (19)$$

In my own practice, the sample-size n/N is nearly always so small that the 2d term is negligible.

3. The term $h^2/n_1 n_2$ is the approximate relative increase in Var X that arises from failure to make an exact Neyman allocation. Or, it is the relative increase in the sample-size n that is necessary to restore the Var X to what it would have been for exact Neyman allocation. We shall see later in the numerical illustrations how easy it is to use this term as a guide.

4. We may now plot Var X against n_1. The curve is a parabola, vertex down; it is very flat in the neighborhood of the vertex. Hence ANY REASONABLE APPROXIMATION TO THE NEYMAN ALLOCATION WILL GIVE EXCELLENT RESULTS. This is not to say that any allocation will be good, but that an honest, sincere attempt, based on some previous knowledge of σ_1 and σ_2 may give us results almost as good as the exact values would give.

We may generalize Eq. 16 for any number of strata by setting

$$n_i = \frac{P_i \sigma_i}{\bar{\sigma}_w} + h_i$$

where $\Sigma h_i = 0$. The result is

$$\text{Var } \bar{x} = \frac{(\bar{\sigma}_w)^2}{n} \{1 + \Sigma \frac{h_i}{n} \frac{h_i}{n_i}\} - \frac{\sigma_w^2}{N} \tag{20}$$

If the numbers h_i are all small, Var $\bar{x}$ will be but little bigger than if the allocations were Neyman exactly.

One may alter the equations for Var X to show the variance of $\bar{x}$ (merely divide through by N^2). A summary of results is below.

A: $$\text{Var } \bar{x} = P_1^2 (1 - \frac{n_1}{N_1}) \frac{\sigma_1^2}{n_1} + P_2^2 (1 - \frac{n_2}{N_2}) \frac{\sigma_2^2}{n_2} \tag{21}$$

[Plan A, any allocation with the sample-sizes fixed in advance]

B: $$\text{Var } \bar{x} = (1 - \frac{n}{N}) \frac{\sigma_w^2}{n} \tag{22}$$

[Plan B, proportionate allocation, sample-sizes fixed in advance]

C: $$\text{Var } \bar{x} = \frac{(\bar{\sigma}_w)^2}{n} - \frac{\sigma_w^2}{N} \tag{23}$$

[Plan C, Neyman allocation, sample-sizes fixed in advance]

<u>What do we lose by small departures from Neyman allocation</u>? The answer is that we lose very little by <u>small</u> departures. The theory is contained in Eq. 16 for 2 strata, and in Eq. 20 for more strata. In an actual example, I was fairly sure that the ratio $P_1\sigma_1 : P_2\sigma_2$ was about 46:54, wherefore by Eq. 17, the ratio $n_1 : n_2$ should be 46 : 54.

This is a difficult ratio to work with, and I hesitated to prescribe it: I decided to make $n_1 = n_2$. How much precision did this decision cost? It may be obvious that my decision was equivalent to setting

$$\frac{h}{n_1} = \frac{4}{46}, \qquad \frac{h}{n_2} = \frac{4}{54}$$

whence

$$\frac{h}{n_1}\frac{h}{n_2} = \frac{4}{46}\frac{4}{54} = .08^2 \tag{24}$$

Then the bracketed term in Eq. 20 will be

$$1 + .08^2 = 1.0064$$

which indicates a loss of only 6 interviews in 1000--far too trivial to mention, and far within the limits of the uncertainty in the advance knowledge of $P_1\sigma_1 : P_2\sigma_2$. The decision was a wise one, but I did not feel safe in it until I saw this delightful numerical result.

Comparisons of variances obtained by using no stratification, proportionate allocation, and Neyman allocation. We have before us on previous pages the variances of these three plans, A, B, and C. Comparison is only a matter of algebra, and as an exercise the reader may show that if the total sample (n) is the same in all three plans, then

$$\left.\begin{aligned} \frac{A-B}{A} &= 1 - \left(\frac{\sigma_w}{\sigma}\right)^2 = \left(\frac{\sigma_b}{\sigma}\right)^2 \\ \text{or}\quad B &= A\left(\frac{\sigma_w}{\sigma}\right)^2 \end{aligned}\right\} \quad \begin{bmatrix}\text{The relative gain of} \\ \text{proportionate allo-} \\ \text{cation over no} \\ \text{stratification}\end{bmatrix} \tag{25}$$

$$\left.\begin{aligned} \frac{B-C}{B} &= 1 - \left(\frac{\bar{\sigma}_w}{\sigma_w}\right)^2 \\ \text{or}\quad C &= B\left(\frac{\bar{\sigma}_w}{\sigma_w}\right)^2 \\ &= A\left(\frac{\bar{\sigma}_w}{\sigma}\right)^2 \end{aligned}\right\} \quad \begin{bmatrix}\text{The relative gain of} \\ \text{Neyman allocation} \\ \text{over proportionate} \\ \text{allocation}\end{bmatrix} \tag{26}$$

These are very important equations. They will tell us whether protionate or Neyman sampling in designated strata will show a gain in precision over unstratified sampling, and how much, provided we know something about σ_w/σ, or about $\bar{\sigma}_w/\sigma_w$.

In my own practice, I usually make calculations based on two strata. If two strata show no appreciable gain, then there is no use to try three. But if two strata show some appreciable gain, then three or more strata, carefully defined, may show a further gain.

Some simple numerical illustrations. In order to see some numerical results, suppose that we are going to take a sample over a region to discover the total number of readers of a particular magazine. The mailing list and the number of copies sold by dealers enables us to divide the area into two parts such that

$p_1 = .10$, the proportion of readers in Stratum 1

$p_2 = .01$, the proportion of readers in Stratum 2

$P_1 = .5$, the proportion of sampling units in Stratum 1

$P_2 = .5$, the proportion of sampling units in Stratum 2

Then

$$\sigma_1 = \sqrt{p_1 q_1} = .30 \qquad \sigma_2 = \sqrt{p_2 q_2} = .10$$

$$\sigma_w = P_1\sigma_1 + P_2\sigma_2 = .20 \qquad (\bar{\sigma}_w)^2 = .04$$

$$\sigma_w^2 = P_1\sigma_1^2 + P_2\sigma_2^2 = .05$$

$$p = P_1 p_1 + P_2 p_2 = .055, \text{ the overall proportion of readers}$$

$$\sigma_b^2 = P_1(p_1 - p)^2 + P_2(p_2 - p)^2 = .002$$

$$\sigma^2 = \sigma_w^2 + \sigma_b^2 = .052$$

Let us compare the three variances A, B, C. First, to compare Plans A and B we use Eq. 25 and see that

$$B = A\left(\frac{\sigma_w}{\sigma}\right)^2 = A\,\frac{50}{52} = .96\ A \tag{27}$$

Thus, 96 interviews by proportionate sampling would give us the same precision as 100 interviews unstratified. The gain is small, and I should recommend proportionate stratified sampling only if the cost and effort of stratification were practically negligible.

Now let us see if Neyman allocation will be better. We use Eq. 26 and see that

$$C = B\left(\frac{\bar{\sigma}_w}{\sigma_w}\right)^2 = B\,\frac{.04}{.05} = .80\ B \tag{28}$$

Thus, 80 interviews by Neyman sampling will give us the same precision as 100 interviews allocated proportionately. In this case, Neyman allocation would probably show a net gain over its additional cost.

Remark. One must be careful not to generalize from one illustration. Sometimes the difference B - C between

proportionate and Neyman sampling will be negligible. Sometimes neither will show any appreciable gain over an unstratified sample. One must depend on theory, not hunches. Fortunately, the calculations require only a few minutes.

Modification when the costs vary greatly from stratum to stratum. In the foregoing pages there was an assumption that the cost of an interview (or of a test) is the same in one stratum as in another. Sometimes, however, the costs will be so greatly different that it will be a good idea to modify the allocation of the sample to the strata. The solution is very simple--decrease the size of the sample in any stratum where the costs are relatively excessive, and increase the sample where the costs are relatively very cheap, keeping the total cost the same as it would have been otherwise. This solution applies to any of the plans in this paper.

Specifically, for example, in place of the straight Neyman allocation (Plan C) we may use now the allocation

$$n_i : n_j = \frac{N_i \sigma_i}{\sqrt{c_i}} : \frac{N_j \sigma_j}{\sqrt{c_j}} \qquad [\text{Plan C}'] \tag{29}$$

which we shall call Plan C'. Because the costs enter the equation only under the root-sign, it will pay to replace Plan C by Plan C' only when the cost in one stratum is considerably greater than the cost in another stratum. A very simple calculation will show whether under a given set of costs and variances Plan C' will show much saving over Plan C.

One point to remember is that if one is going to use Neyman allocation (Plan C) anyway, he will cause himself but little further inconvenience and cost by introducing Plan C', so that even if the probable saving is only 5 or 10%, one might as well have it in his pocket.

An example will help. Suppose that a survey is to cover a region that consists of an urban area (Stratum 1) where the cost of an interview averages \$5; also the surrounding rural area (Stratum 2) where the cost is \$10. Suppose that the proportions of sampling units in the two areas are 40% and 60% of the total, and that the variances between sampling units (for some particular characteristic) are in the ratio 2 : 1. Then

$$P_1 : P_2 = .6 : .4$$

$$\sigma_1 : \sigma_2 = \sqrt{2} : 1$$

and Plan C' gives

$$n_1 : n_2 = P_1\sigma_1 \sqrt{c_2} : P_2\sigma_2 \sqrt{c_1}$$

$$= 6 \sqrt{2} \times \sqrt{10} : 4 \times 1 \times \sqrt{5}$$

$$= 12 : 4 = 3 : 1$$

whereas Plan C would give

$$n_1 : n_2 = P_1\sigma_1 : P_2\sigma_2$$

$$= 8.5 : 4 \quad \text{or} \quad 17 : 8$$

Now suppose that we have $2000 to spend on the field-work. How will the variances of the two plans C and C' compare?

Plan C':

$$5n_1 + 10n_2 = 2000$$

$$5n_1 + 3.33n_1 = 2000$$

$$n_1 = 240$$

$$n_2 = 80$$

$$n = 320$$

$$\text{Var } x = P_1^2 \frac{\sigma_1^2}{n_1} + P_2^2 \frac{\sigma_2^2}{n_2}$$

$$= .6^2 \frac{2}{240} + .4^2 \frac{1}{80} = .00500$$

Plan C:

$$5n_1 + 10n_1 = 2000$$

$$5n_1 + 40n_1/8.5 = 2000$$

$$n_1 = 206$$

$$n_2 = 97$$

$$n = 303$$

$$\text{Var } x = \frac{(P_1\sigma_1 + P_2\sigma_2)^2}{N}$$

$$= \frac{(.85 + .4)^2}{303} = .00515$$

Plan C' gives thus only 3% lower variance than Plan C, and is hardly worth while.

For further comparison we may see what happens with proportionate allocation (Plan B), in which $n_1 : n_2 = 6 : 4$.

$$5n_1 + 10n_2 = 2000$$

$$5n_1 + 40n_1/6 = 2000$$

$$n_1 = 171$$

$$n_2 = 114$$

$$n = 285$$

$$\text{Var } \bar{x} = \frac{P_1\sigma_1^{\,2} + P_2\sigma_2^{\,2}}{n}$$

$$= \frac{.6 \times 2 + .4 \times 1}{285} = .00598$$

This example happens to be one in which the Neyman allocation shows a good gain over proportionate allocation, and in which further adjustment for costs (Plan C') accomplishes little more.

One must be careful not to generalize, but to treat each example by itself, on the basis of theory and the best information available concerning costs and variances.

PART D. STRATIFICATION AFTER SELECTION

Why stratify the entire frame before we draw the sample? It is not necessary to classify all the sampling units in the entire frame before we draw the sample. When the frame contains thousands of sampling units, the cost of classifying every unit may be prohibitive, and it may be preferable to classify only a sample of sampling units, and in this way to decrease the cost and speed up the work.

There are two types of procedures by which to dodge the classification of every unit in the frame. One type corresponds to proportionate allocation (Plans D and E) and the other corresponds to Neyman allocation (Plans F and G). It is simplest to think of these plans in terms of a preliminary sample which we draw without stratification just as if it were Plan A. The preliminary sample then becomes a miniature frame to which we apply procedures already learned, with some variations. In Plans D, E, and G we require advance knowledge of the weights P_i (or of the numbers N_i) and also information in advance by which to classify a sampling unit into one stratum or another: in Plan F we do not; there we derive this information, or part of it, from preliminary interviews. We shall write out the plans by steps.

PLAN D (Weights P_i known)

1. Draw a sample of n sampling units from the entire frame without stratification (as in Plan A).

2. Classify the n sampling units into strata.

3. Carry out the interviews or the tests on the entire sample of size n.

4. Calculate the separate estimates

$$X_1 = N_1\bar{x}_1, \qquad X_2 = N_2\bar{x}_2, \qquad \text{etc.} \tag{30}$$

for the populations stratum by stratum, using the known values of N_1, N_2, etc., but with the mean populations $\bar{x}_1$, $\bar{x}_2$, etc., formed from the samples.

5. Consolidate these separate estimates to form the estimate $X = X_1 + X_2 +$ etc., of the total population A of the entire frame.

6. Estimate the variance of $\bar{x}$. The easiest way is to lay out the sample in the first place by the Tukey plan, but with more labor one may use the formula below, which one uses also in the planning:

$$\left.\begin{aligned} &\text{Var } \bar{x} = \underbrace{\frac{1}{n}\{(1 - \frac{n}{N})\sigma_w^2}_{\text{Plan B}} + \frac{1}{n}\sigma_R^2\} \quad [\text{Plan D}] \\ &\sigma_R^2 = Q_1\sigma_1^2 + Q_2\sigma_2^2 + \text{etc.} \quad \left[\begin{array}{l}\text{The reverse}\\ \text{internal}\\ \text{variance}\end{array}\right] \\ &Q_i = 1 - P_i \end{aligned}\right\} \tag{31}$$

PLAN E (Weights P_i known)

1. Fix the sample-sizes n_i as in Plan B (called quotas hereafter).

The term "quota" used here bears no relation to the use of the same word for a selection by the interviewer, a non-probability method that I do not use.

2. Draw one by one sampling units from the frame, without stratification, as in Plan A, and classify each unit into a stratum as you draw it. (Draw groups of 5 or 10 units at a time if you prefer.) Continue until the quotas n_1, n_2, etc. are all exactly filled. In doing so, reject any sampling unit that belongs to a stratum whose quota is already filled.

3. Carry out the interviews or the tests on the final sample.

4. Form the estimate

$$X = \frac{N}{n} S = N\bar{x} \tag{32}$$

exactly as in Plan B. X is an unbiased estimate of A.

5. Estimate the variance of X or of $\bar{x}$ exactly as you would in Plan B.

Remark. In Plan E the sample-sizes are fixed in advance: in Plan D they are not; they are random variables.

Choice between Plans D and E. It was presupposed in the treatment of both these plans that information exists in records that are already on hand or obtainable (as by purchase of a directory or of Census tables) by which to classify a sampling unit into one stratum or another. In respect to costs, they are about equal for a prescribed precision. If the tabulations are simple, and if there is little extra weighting to do in the formation of the estimates, then there will be little difference between the two plans. The simplicity of the self-weighted estimates of Plan E in Eq. 32 may then be a deciding factor: otherwise the choice may well rest on the basis of personal preference.

But this is not the whole story. Both Plans D and E will often be used to obtain estimates of proportions in fine classes, such as the proportion of males that are of age 20-29 and employed in a particular occupation; also for ratio-estimates of the total population in such classes. In a heavy tabulation program, Plan E may well possess distinct advantages, because of its self-weighting feature, expecially if there is little extra weighting to do.

Neyman allocation of the preliminary sample (Plan F). In Plans D and E the sample-sizes n_1, n_2, etc. were nearly or exactly proportionate. In Plans F and G we shall adjust them to the Neyman allocation.

PLAN F
(Weights P_i not known)

1. Decide on the most likely ratios $\sigma_1 : \sigma_2 : \sigma_3$ for the chief characteristic that the sample is expected to measure. These ratios fix the final ratios $n_i : N_i'$ by the Neyman relations

$$\frac{n_1}{N_2'} : \frac{n_2}{N_2'} : \frac{n_3}{N_3'} = \sigma_1 : \sigma_2 : \sigma_3 \qquad (33)$$

which comes from Eq. 18 . N_1', N_2', etc. are the sizes of the classes in the preliminary sample of total size N' (next step).

2. Compute the optimum size N' of the preliminary sample, and draw it from the frame, without stratification, as in Plan A. The optimum size for the preliminary sample will be seen in Eq. 38.

3. Classify each of the N' units into its proper stratum. This will require a short study of each sampling unit--perhaps a study of previous census information, or a study of the files or other records, perhaps a brief interview or a quick test to determine which stratum it belongs to.

4. Reduce the number of units in each stratum to reach the final ratios as given by Eq. 33 and to reach also the final total sample-size n.

5. Form the estimates $\bar{x}_1$, $\bar{x}_2$, etc., and then of $\bar{x}$.

6. Estimate the variance of $\bar{x}$. The easiest way is to lay out the sample in the first place by the Tukey plan, but with more labor one may use the formula

$$\text{Var } \bar{x} = \frac{(\bar{\sigma}_w)^2}{n} + \frac{\sigma_b^2}{N'} \tag{34}$$

This is also the formula that one uses in the planning stages, to decide the total sample-size n.

PLAN G
(Weights P_i known)

1. Fix the sample-sizes n_i by Eq. 18. This is possible because we know the weights P_i, and presumably also the standard deviations σ_i.

2. The same as Step 2 in Plan E.

3. The same as Step 3 in Plan E.

4. Form the estimate $\bar{x}$ as in Plan C.

5. Estimate the Var $\bar{x}$ by Eq. 23 as in Plan C.

<u>Application of Plan F to determine the condition of the aerial plant of a telephone company</u>. The use of Plan F will sometimes bring forth considerable saving in the sampling of materials. One example is the selection of items of the aerial plant of a telephone company, where the aim of the survey is to estimate the per cent condition of the aerial plant. The purpose of the sample is to determine the physical depreciation of the various kinds of items that constitute the aerial plant of the company. Aerial cable is usually a very valuable part of the plant; yet perhaps only one pole in 4 (see figures further on) carries cable. The other poles carry aerial wire, which is not so valuable.

Expert inspectors will examine each pole in the sample, plus the other aerial plant attached thereto, and will record the physical condition (new or good as new, slightly used, etc.) of each type of item (pole, aerial cable, copper wire, iron wire, cross arms, etc.).

Plan F is especially useful when a small proportion of the poles carry aerial cable, <u>and</u> when the aerial cable forms a substantial portion of the value of the aerial plant. It is then possible to concentrate a large portion of the total dollar-value of the plant into one recognizable type of pole (Class 1 below). The same procedures are applicable equally to the sampling of underground plant, especially if a small proportion of the manholes contain a large fraction of the total underground plant. The inspections of the aerial plant takes place on poles: the inspection of the underground plant takes place in manholes.

The procedure in the case of aerial plant is to divide the poles on the record into 2 classes:

Class 1. Poles that according to the engineering records carry aerial cable.

Class 2. All other poles.

In the case of the underground plant, the subdivision would be

Class 1. Large manholes (e. g., those that contain 18 or more ducts on all sides).

Class 2. Smaller manholes (those with fewer ducts).

By a development similar to the deriviation of the Neyman allocation, it is fairly simple to prove that if $\sigma_1 = \sigma_2$ (an assumption that experience shows is good enough), the optimum allocation of the efforts of the inspectors will be obtained if the sample-sizes in the two classes are in proportion to the weights of the classes, so that

$$\left.\begin{aligned} n_1 &= w_1 n \\ n_2 &= w_2 n \end{aligned}\right\} \qquad (35)$$

The weights w_1 and w_2 are the proportions of the dollars in the two classes (supposed known).

The first question is which plan to use: A, D, or F? Plans B and C are impracticable because there are 600,000 poles--too many to classify in advance. Plans E and G are impossible because we know not exactly the proportions P_1 and P_2.

Suppose that we investigate the relative efficiency of Plan F over Plan A, to get an idea of which one of these two to choose.

Suppose that the accounting department is able to give us figures for w_1 and w_2, viz., that

$$w_1 : w_2 = 7 : 3$$

Suppose further that the engineering department has an approximate figure of 25% for the proportion P_1 of the poles that carry aerial cable. This figure gives

$$P_1 : P_2 = 25 : 75$$

Previous experience shows that one may expect σ_1, σ_2, and σ_3 to be about 12%, and that σ_b may be anywhere from 1 to 2%: as an approximation we set $\sigma_b = 1.5\%$. As for the costs c_1 and c_2, I had learned that a girl that earns $20 per day can classify about 100 poles per day, wherefore c_1 is about 20¢ per pole. A pair of inspectors, with a truck and tools, can inspect about 10 poles per day, wherefore c_2 is about $10 per pole.

We next observe that we may ignore the term σ_b^2/N' in Eq. 34 for the variance by Plan F: this is so because σ_b is small and because we

know already or shall soon see that N' (our preliminary sample) will be large, probably in the neighborhood of 1000 or more (actually 2240; see the table infra). Hence, for the proportionate efficiency of Plan F over Plan A we have in this case the simple equation

$$\frac{A}{F} = \frac{w_1^2}{P_1} + \frac{w_2^2}{P_2} \tag{36}$$

which gives

$$\frac{A}{F} = \frac{.7^2}{.25} + \frac{.3^2}{.75}$$

$$= 1.96 + .12$$

$$= 208 : 100 \tag{37}$$

Seeing this numerical result, we instantly adopt Plan F, because 100 inspections carried out by Plan F will be equivalent to 208 by Plan A, wherein we should simply draw poles by random numbers and inspect them as they come.

Now comes the question of the sizes of the samples. First, the preliminary sample N'. The formula for the optimum ratio n : N' is

$$\frac{n}{N'} = \frac{\bar{\sigma}_w}{\sigma_b} \sqrt{\frac{c_1}{c_2}} \tag{38}$$

which gives

$$\frac{n}{N'} = \frac{15}{1.5} \sqrt{\frac{20}{1000}} > 1 \tag{39}$$

As this ratio is greater than 1, our preliminary sample need only be big enough to supply the required number of poles in Class 1.

Suppose that we desire $\sigma_{\bar{x}}$ to be about .30 per cent. Then, if $\sigma = 12\%$ Plan A would require the inspection of

$$n = \left(\frac{\sigma}{\sigma_{\bar{x}}}\right)^2$$

$$= \left(\frac{12}{.30}\right)^2 = 1600 \text{ poles} \tag{40}$$

The required final sample by Plan F will be about half this number (more strictly 100:208). By a bit of arithmetic we are able to draw up

the accompanying table and to write down the following steps for selection:

1. Select by random numbers a preliminary sample of poles;
2. Determine from the engineering records which poles carry cable;
3. Retain for inspection all the poles that carry aerial cable.
4. Retain for inspection 1 pole at random from every successive 7 that do not carry aerial cable.

The ratio of 1:7 for the thinning is correct because it produces final sample-sizes that have the required ratio 7:3.

Class	Preliminary sample	Final sample
Both classes	2240	800
With cable	560	560
Without cable	1680	240

The final sample of 800 poles is equivalent to 1600 by Plan A. The saving of Plan F over Plan A is

$$(1600 - 800)\ \$10 - 2240 \times \$.20 = \$7552$$

On some jobs the gain will not be so great. The gain came in this instance from the fact that w_1 was large and P_1 was small. If P_1 were larger, the gain would be less. Thus, if 33% of the poles carried aerial cable, the relative efficiency of Plan F over Plan A would be

$$\frac{A}{F} = \frac{.7^2}{.33} + \frac{.3^2}{.67} \qquad [\text{Eq. } 36]$$

$$= 160 : 100 \qquad (41)$$

This is still a sizable gain, but a big drop below the former gain of 208:100.

As the gain obviously falls off sharply with an increase in P_1, one must be prepared to accept some loss in precision from the fact that the prior information on the number of poles that carry aerial cable may be in error, and that the thinning ratio prescribed may consequently not produce the required precision. In anticipation, it is wise, in the absence of firm information, to specify samples a bit bigger than theory indicates.

ADVANTAGES AND APPLICATIONS OF STATISTICAL QUALITY CONTROL IN THE AIRFRAME INDUSTRY

J. L. Coburn
Convair, Fort Worth Division

"Statistical Quality Control certainly looks like it has a great deal of merit, but I'm afraid it has little or no place in this business because we are so different from other types of industry. We have enough paperwork now, without adding the complications which appear to be connected with this type of thing." Were it possible to line up, head to toe, all of the statistical quality control engineers who have experienced this statement, or who have come to grips with the convictions that generated it, we would probably have an unbroken line stretching from Chicago to the ASQC offices in New York City.

In regard to the production of aircraft, the frequency of its occurrence is such that the initial task of the quality control engineer is well defined even before he formulates the details of his program. Failure to recognize the negative impact that this type of thinking can ultimately have on a statistical quality control installation has been responsible for the deaths of many basically sound programs. Worse yet for the profession as a whole, were those that did not succumb completely but which assumed complete anonymity or were aborted to token programs, with little or no purpose other than to create a surface impression for visiting firemen, lull management into a false sense of security and to provide a steady income for the statistician. The latter condition generally occurs when responsible supervision is reluctant to admit that the original objective has been compromised.

As we all know, Statistical Quality Control has been in the wings of the American industrial scene for a good many years. It made its initial appearance in the aircraft industry during the war years, when management was casting about in desperation for a pill which would cure the ill of production rejects - the unit of production that had to be sent back to the machine operator for additional work, or the piece that had to be thrown away because it had been ruined by a careless operator. Further, a tool was needed to reduce inspection time, yet provide the necessary assurance that the outgoing quality level would be maintained. At about this time, well-qualified and conscientious scientists produced a very effective but extremely delicate tool with which the reject disease might be treated. Many programs were immediately set into motion. Some failed and some were highly successful. Those which succeeded owe their effectiveness to sound programming, maximum utilization of the psychological factors which buttress any good statistical quality control program; and last, but by no means least, by effecting a union between quality, cost, and schedules. The failures, for the most part, may be traced to one thing: Too much emphasis on statistics, and too little on what the statistics were supposed to do.

It is not my purpose to expound at any great length on the philosophical aspects of why certain programs failed or succeeded. Progression from the war years has enabled all of us to review what we did wrong. The important thing, of course, is to profit from what we have learned, to the end that we have a clear recognition of the task which confronts us today. The word 'today', as used here, is synonymous with the prevailing austerity concept which has been adopted by the Government; particularly, where military procurement is concerned. Further, 'reliability' is a word that is no longer a mystic connotation which can

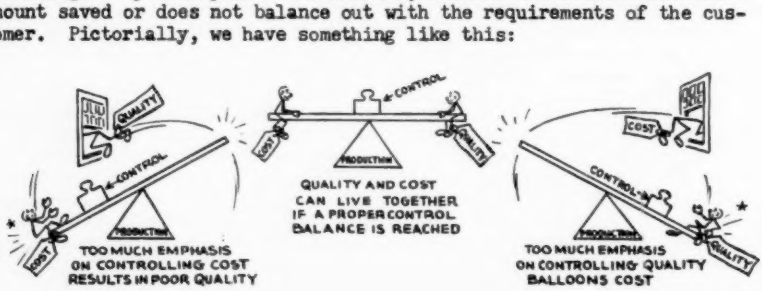
be seen on the horizon. Both of these things are here with us today. Recent trends indicate a slackening off of the 'cost plus-fixed fee' policy. The 'fixed price' concept is certainly not some nebulous thing which will all be taken care of in the carpeted offices of top management. Its effect is already being experienced in the management, tooling, and purchasing areas, and eventually will have its impact on the machine operator himself. Each minute of rework, each portion of a standard hour lost through scrappage, will bring about a reduction in corporation profits. In addition to this, if rework and scrap costs are nebulous and indistinct, the estimating section will be seriously handicapped in their efforts to provide for these overhead contingencies in developing figures for competitive bidding. However, merely reporting accurate scrap and rework factors is not enough. A supplementary program must be developed whereby these costs may be controlled on the operating level. To be controlled, these cost factors must be identified and made important to each and every member of the organization who designs, plans, transports, performs operations on, or inspects manufactured items.

Today, in the field of aircraft, there are as many variations of statistical quality control programs as there are companies. Each has been tailored to fit the needs of the particular organization within which it functions. The main purpose of this paper is to outline in general some of the activities of Convair's Statistical Quality Control Program, and to present the method which we intend to use relative to controlling rework and scrap costs.

WHAT IS STATISTICAL QUALITY CONTROL?

Quality control is a concept sufficiently resilient to lend itself to several practical and reasonable definitions. A. V. Feigenbaum, in his book, "Quality Control, Principles, Practice and Administration", defines it as "an effective system for coordinating the quality maintenance and quality improvement efforts of the various groups in an organization so as to enable production at the most economical levels which allow for full customer satisfaction".

At Convair, we have developed another definition which does not deviate to any great extent from the above, but which in our estimation, meets the requirements of our own program. The definition is based on the axiom which states that "progress is indeterminate unless it can be measured", and further, on the premise that quality follows the law of diminishing returns. In other words, reduction of discrepancies results in savings - up to a point - thereafter, control costs more than the amount saved or does not balance out with the requirements of the customer. Pictorially, we have something like this:



To attain full stature within the organization therefore, Quality Control must pull its own weight and pay some of its overhead costs by developing a control facility, whereby quality and cost may be integrated and interpreted from the operating level though top management. Our definition, therefore, is: "Statistical Quality Control is an effective tool for measuring and controlling the economiquality efforts of the organization and of its segments in order to determine where we are and where we want to go".

WHAT IS EXPECTED OF IT?

Very simply stated, oftentimes much more is expected than it can produce, in itself. The mere existence of organized and well-presented data has never solved a single problem. The information must be studied, digested, and intelligently used, not only by management, but by all successive levels. Specifically, however, our experience indicates that the following listed characteristics are generally required of a statistical quality control section, by management:

A. Existing Production Programs
   1. Collection, organization and tabulation of quality data, leading to development of clear, timely and concise reports to all levels relative to rejection, rework, and scrap position (Inplant production and Outside-procured materials)
   2. Development and administration of sampling techniques
   3. As required, special process controls (in the case of aircraft production where lot sizes are not consistently large, these controls are generally restricted to certain expensive large lot items)
   4. Quality incentive programs (recognition of improvement, quality leaders, etc.)
   5. Chronic discrepancy control

B. New or future Production Programs
   1. Statistical research on newly-developed processes
   2. Machine capability studies
   3. Test data correlation
   4. Tolerance studies (generally in conjunction with Engineering)

C. Special Projects
   1. Development of programs, such as controlling and predicting reliability levels of complex electrical or electronic systems.
   2. Special studies where application of the statistical science can implement the required information.

Obviously, time does not permit a complete description of the abovementioned activities. Each in its own way is equally significant, but the one which quickly lends itself to an explanation of the advantages of a statistical quality control system is the reporting and control of rejections, rework and scrap. Consequently, the remainder of this presentation will be in that direction.

REQUIREMENTS OF A REWORK AND SCRAP CONTROL PROGRAM

A. Total Measurement of Direct Labor Expended in Rework
   Quantitative figures representing number of pieces rejected as requiring rework are not sufficient for ultimate control. Per cent defective is much more useful if it can be interpreted in

terms of direct labor lost. A department may be operating at a very low percent defective, and at the same time suffering tremendous rework costs. Rarely can these costs be identified for corrective action by efficiency reports alone.

B. Total Measurement of Scrap Costs
Again quantitative measurement of pieces scrapped is not adequate as a tool for predicating corrective action. Standard hours scrapped converted to direct labor hours lost, plus the material costs involved, are a much more effective measure of performance.

C. Chronic Discrepancy Control
The bogey man in any industrial activity is represented by unnecessary repetitive costs. Those items which were unacceptable to begin with but through failure to maintain proper control that they might be quickly recognized and eliminated, a felony was compounded, resulting in multiplication of operating costs and subtraction of profits.

D. Correlation of Quality and Cost
Of course it is necessary to meet the requirement of the customer - if you don't, you're out of business. However, it is not good practice to maintain rigid control over the color of a product if the customer is more concerned with the smell of it. Somewhere in a program, if the standards are not consistent or fluctuate wildly at the whim of inspection, we begin to pay heavily for an unnecessary luxury. The by-products of inconsistent standards are the confusion which is visited upon the Production departments and the loss of respect for inspection, all serving to balloon costs.

E. Practical, Economical, and Hard-hitting
The law of diminishing returns requires that the control section be tailored and staffed to achieve just what is required, and nothing more. Proper organization and full utilization of mechanical data processing machinery can assure this. Sufficient time must be provided for expert analysis and work on the floor level. Reports and visual controls must be simple, graphic wherever possible, and not cluttered up with extraneous gobble-dygook that has no significance. Each level, from the machine operator or mechanic to top management, must be approached on their own terms. The machine operator or mechanic is interested only in how he himself is doing, not Henry Jones in another cost center of the department. On the other hand, the general foreman is interested in how each cost center or station is doing, and what they are individually contributing to the operation of his department. Finally, management wants to see the overall picture, both in terms of quality levels and cost. Nothing will torpedo a statistical program as fast as placing reams of reports which must be carefully correlated and analyzed before a conclusion can be reached, on the desk of any level of supervision or management, where time for making decisions is at a premium.

F. Cement Customer Relationships
It is generally conceded, particularly where military procurement is concerned, that the customer is interested in seeing

what manner of control is exercised over the methods used to spend his money or, at the very least, to see where the money is going. A good sound "snow-job-less" rejection and scrap control program can alleviate a great deal of distress when it comes to dealing with the customer. In view of manpower limitations, the customer does not always have available sufficient data upon which to base a conclusive and objective review. He must therefore, when evaluating product quality trends, use statistical data which has been developed by the manufacturer. Should this data be inconclusive or subject to error, a great deal of needless wrangling is experienced by both parties. This, of course, does not benefit mutual relations. Thus, it behooves the prime contractor to develop factual and realistic data which can be used mutually by himself and the customer.

BASIC FOUNDATION OF SYSTEM

A. Classification of Discrepancies
The entire system is predicated upon the fact that errors will exist in any production activity. To assure a standard procedure, the error should be defined and categorized for the inspector, as to its relation to the end product. The definitions which we have chosen to use are closely related to those generally applied by the industry; however, they have been altered to better suit Convair's application. The following three classifications, and their relative significance, are submitted as examples:

	CLASS	DEFINITION
I.	Critical	A defect which could result in hazardous or unsafe flight conditions; which could prevent performance of a tactical mission; or which could affect aircraft weight (safety, performance, weight).
II.	Major	A defect other than critical, that materially reduces the usability of the end product; or could cause substantial production difficulty in later stages of manufacture or assembly (interchangeability, service life, assembly).
III.	Minor	A departure from Engineering or Quality standards that has no significant effect on the use or operation of the end product, but which should be reworked or corrected in order to maintain a high level of quality.

B. Standardized Rejection Paper
Two distinctly different types of rejection paper are used. "Critical" and "Major" type discrepancies are processed on the regular Materials Review form called the Inspection Rejection Form, while "Minor" type errors are handled on the Inspection Minor Rework Form. Benefits of this policy are:
(1) Standardization of rejection paper
(2) Segregation of the more serious type discrepancies from

those of a minor nature

(3) Concentration of corrective action on the "Critical" and "Major" discrepancies.

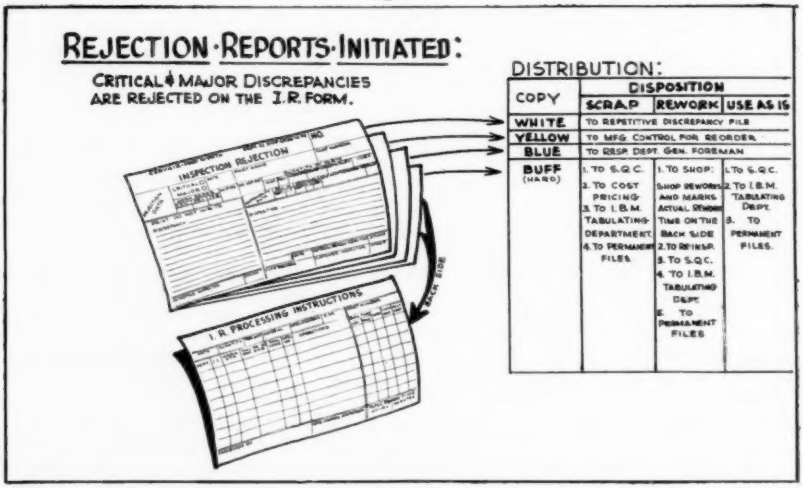
Obviously, more time will be required to process the Materials Review paper (Inspection Rejection Form) through the various scrap pricing and chronic discrepancy control routes, than will be required for processing minor rework data. A survey, run prior to the installation of the new system, indicated that a substantial percentage of minor type discrepancies had been processed on Materials Review rejection paper. By providing a specific form for handling this type of item, two additional benefits were forthcoming:

(1) Due to simplified format, much less time is required for the inspector to fill in the required information.

(2) Reduction in the quantity of Materials Review rejection paper (Critical and Major items) brought about considerable savings in processing time and allowed more corrective action emphasis on the more serious type item.

Figure 1 illustrates the "Inspection Rejection Form" and its processing through the various steps:

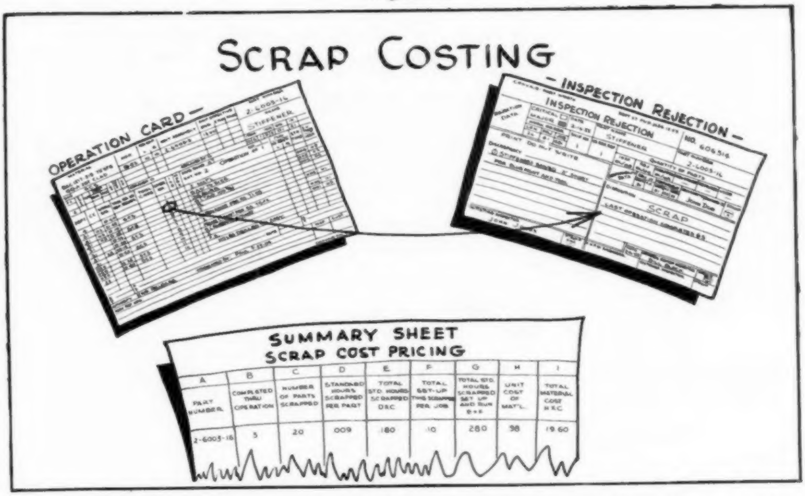
Figure 1



When it is necessary to scrap parts, the Materials Review inspector notes the last operation completed on the face of the Inspection Rejection Form. When the hard copy reaches the Statistical Quality Control section, standard hours scrapped and material costs are calculated and applied to the Inspection Rejection Form. The hard copy then flows to Tabulating. Figure 2 illustrates this activity:

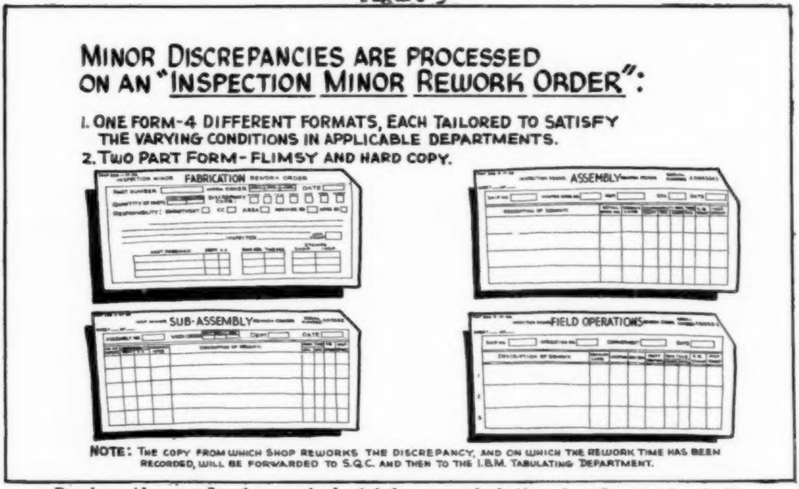
(Paper is continued on the following page)

Figure 2



"Minor" type discrepancies to be processed on IBM-size forms, called "Inspection Minor Rework Orders". This form and its processing are indicated by Figure 3:

Figure 3



During the analysis period which preceded the development of the Inspection Rejection Form and the Inspection Minor Rework Order, it was found that there were present under one roof, four (4) distinctly different manufacturing areas: Fabrication, Sub-Assembly, Major Assembly, and Field Operations. The Inspection Rejection Form (Figure 1) would apply universally to each. It was necessary, however, to tailor the Inspection Minor Rework Order (Figure 2) to each section; thus satisfying the vary-

ing conditions in a pplicable areas. For this reason, the four (4) different formats were developed. This form, by far, experiences the heaviest usage; therefore, it is much less expensive, when revisions are necessary, to change just one of the four forms, rather than all.

INTERMEDIATE PROCESSING OF REWORK AND SCRAP DATA

Figure 4 indicates the daily and weekly IBM summaries which are to be prepared by Accounting and forwarded to the Statistical Quality Control section, for conversion into reports to the various levels of supervision and management:

Figure 4

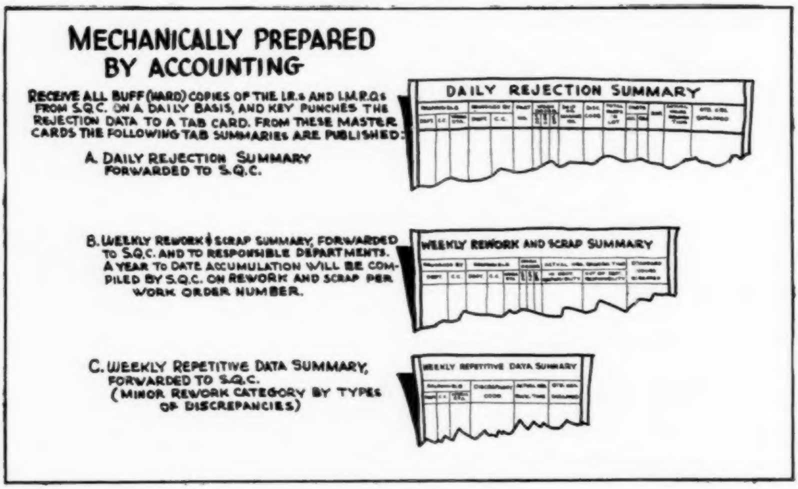
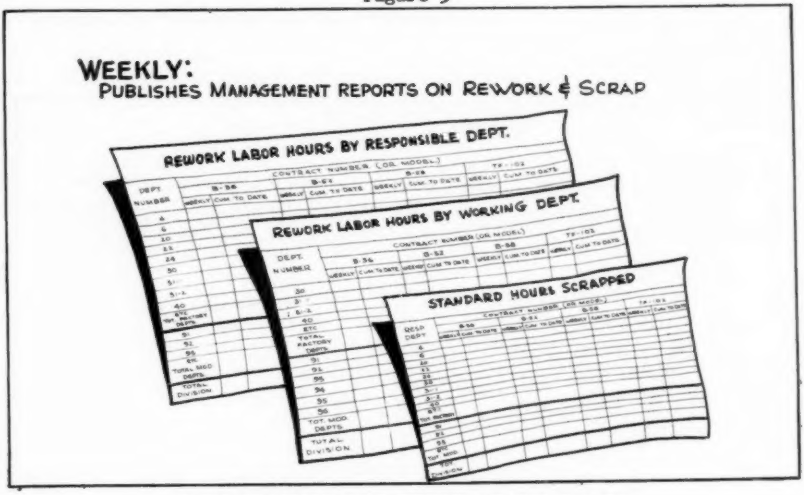


Figure 5 illustrates the first of the "weekly" and "cumulative to date" rework and scrap cost reports. These reports reflect departmental responsibility as previously determined at the floor level:

(Paper is continued on the following page)

Figure 5



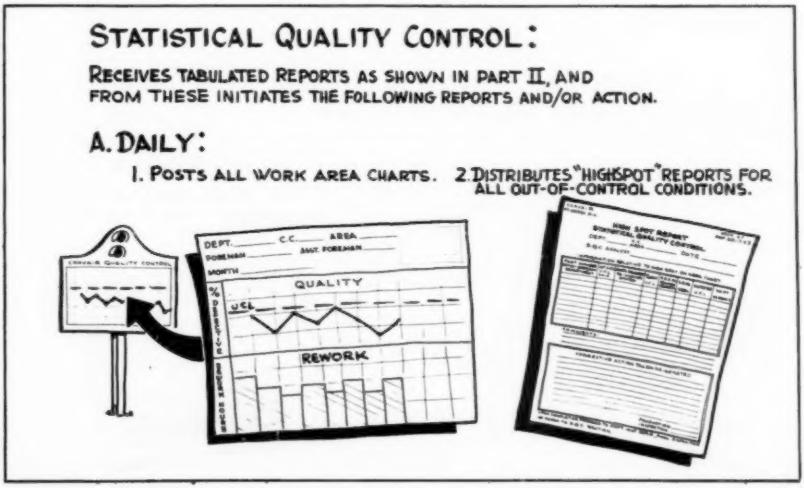
FINAL PROCESSING OF REWORK AND SCRAP DATA

Now that all of the applicable data has been collected, organized, and tabulated, it remains for the Statistical Quality Control section to approach each level with information which is pertinent to that level.

First Level

The employee and the immediate floor supervisor want to see their daily plot on the cost center or station chart (Figure 6):

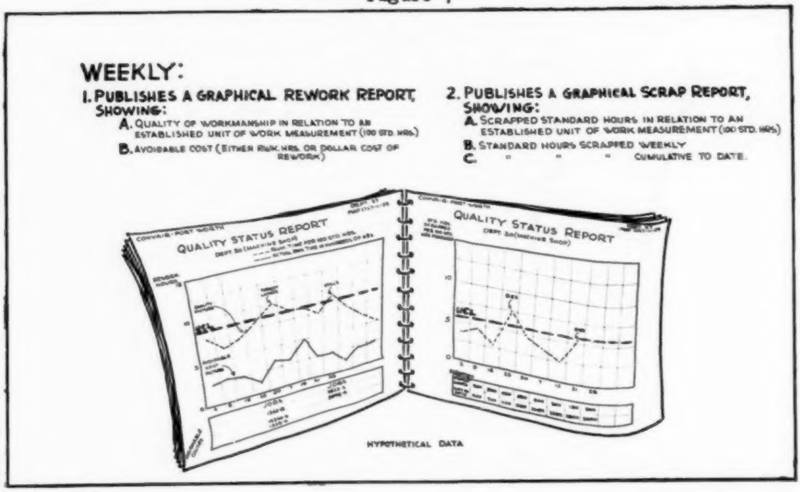
Figure 6



Second Level

The general foreman wants somewhat more comprehensive information concerning his whole department (Figure 7):

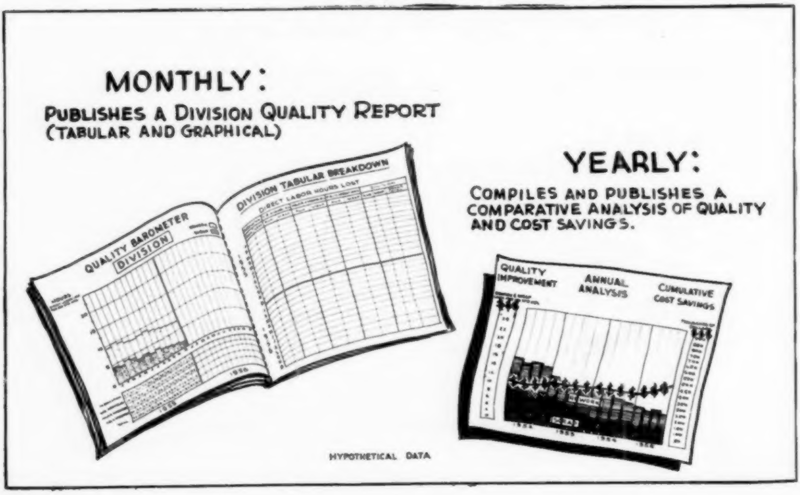
Figure 7



Third Level

Management wants a quick view of the whole forest, plus a tabular breakdown of contrubuting factors. In addition to the monthly management report, a comparative year-to-year analysis of quality levels and cost savings is necessary to button up the entire program (Figure 8):

Figure 8



SUMMARY

The advantages of statistical quality control in the airframe industry are certainly varied and many, and are wholly dependent upon the practicability of the application. It is our opinion that the foregoing rejection, rework, and scrap control is one phase which has a definite place within the organization, if management is at all concerned with constituting itself in order to remain competitive. The current swing toward increased competition demands great emphasis on cost, schedule, and quality. Today, in modern industry where the immediate urgency of war is missing and the competitive element is paramount, poor planning, poor control, and poor quality cannot be tolerated.

A MODIFIED LOT PLOT SAMPLING PROCEDURE FOR CONTROLLING CONTAINER FILL

Leonard Gieseker and LeRoy V. Strasburger*
Field Research Department, National Can Corporation
*Consultant

It is important that close supervision be given to filling operations to see that the maximum possible uniformity is obtained. In the canning of homogeneous products such as pumpkin, the packer is interested only in the net fill weight. In other products, such as canned peas and whole kernel corn, control of not only the net weight but also the drained weight is important. In filling operations, consideration must be given to the change in the drained weight that occurs during processing and subsequent storage.

The majority of fillers used today may be described as volumetric measuring devices. When attempting to fill a definite weight of product through a volumetric measure, certain variables are encountered. Temperature, specific gravity, entrapped air, size and shape of the products and consistency all may introduce filling problems. A number of excellent papers relating to the subject were presented by E. McKinley (1), I. MacPhail (2), H. Link and H. Dobson (3), C. Way (4), H. Edwards (5), W. Brittin (6) at the National Canners Association 1954 Convention.

The majority of vegetables and fruits are canned at the time that they are harvested. The actual canning season covers only a short expanse of time and necessitates numerous temporary employees over this peak period. Under these conditions the problems that confront the quality control personnel are more difficult than when a plant is operated on a continuous basis. Therefore, there is a need for rapid and simple quality control techniques that can be effectively employed. The Lot Plot method developed by Dorian Shainin (7,8) has been extensively used in many industries as an acceptance sampling procedure. It is presented here with the modifications and additions that were found necessary in applying it to filling problems.

Before line control specifications for fill weights can be set up the filler must be evaluated under actual operating conditions. Fill evaluation studies may be used to compare different types of fillers to determine which are the most efficient. They may also be used to test mechanical improvements or different operating conditions of a filler. While the Lot Plot was not designed to be a process control procedure, it can be employed usefully as an indicator of the conditions existing at the time the fill weights are taken. During the relatively short time necessary to run a Lot Plot Fill Evaluation, no assignable causes of extraneous variation should occur in the product. Under such statistictally stable conditions the Lot Plot gives a good estimate of the process chacteristics and may be used, as described in this paper, to establish the standards for line control. Lot Plot Line control is not as accurate a control procedure as control charts for Averages and Ranges. If the expense of the more refined methods can be justified they should be used. Lot Plot control is practical when many of the assignable causes for variation in fill weights are known from past experience. Under these conditions variables found in the daily grading of the canned product are taken into consideration and corrections made before serious difficulty occurs.

A sample size of 50 was selected as a standard. The limitations of

a 50 observation histogram should be recognized. When greater accuracy is desired several Lot Plot Evaluations may be made or a Lot Plot designed for a larger sample size. Anyone having a working acquaintance with statistical methods will recognize the normal distribution curve shown in figure 1. Standard deviation is a measure of the deviation of the observed values (fill weights) from their average. The sign (σ) is used to denote Standard Deviation. If a filler is operating so that the weights found give a normal distribution, then sixty eight percent (68%) of the weights should be found within one standard deviation(1σ) from each side of the average. Two standard deviations will include ninety five and five tenths percent (95.5%) of the weights. Practically all of the weights (99.7%) will be included in three standard deviations from each side of the average. The distance shown as 3σ from each side of the average marks the Lot Limits. Methods of computing Lot Limits for fill distribution that do not produce a normal distribution curve are given by Dorian Shainin (7,8). The Lot Plot is a graphic representation of a distribution that simplifies the calculation of the Standard Deviation. The following paragraphs describe three applications of the Lot Plot method to filling problems.

In some experimental work done for one packer of large dried lima beans the highest, average and lowest fill weights as calculated were filled into cans. This product because of its size and shape is difficult to fill uniformly. The packer was then able to observe the variation that might be expected in any given shipment. During these studies, filler speeds from 155 to 210 cans per minute were tested. An analysis of the results showed that the filler could be increased from 155 to 200 cans per minute with no appreciable increase in variation of fill weights. This resulted in a substantial saving in production costs.

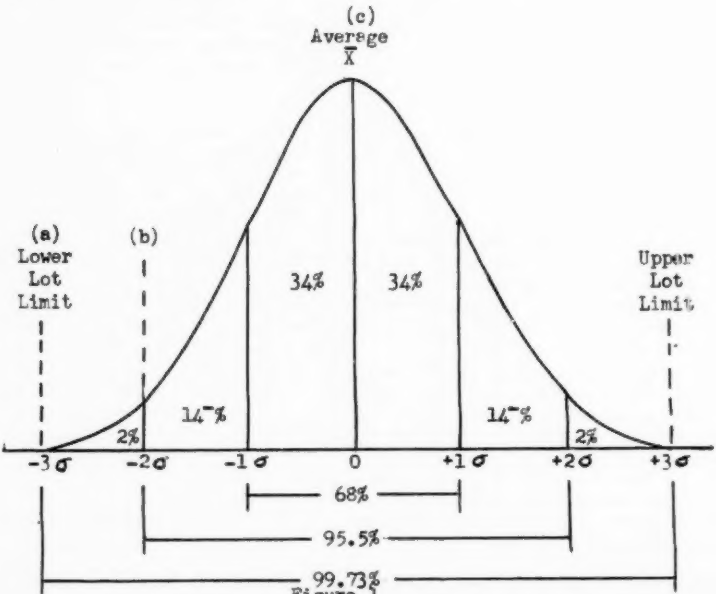


Figure 1

In another instance, the speed of one of the pork and bean fillers was increased and excessive spillage of the sauce occured. A recommendation had been made by the production department to purchase a new filler. The plant management, however, requested a report from the quality control department before this was done. Since it was difficult to actually collect and measure this spillage, Lot Plot Fill Evaluation studies were made to determine the average weight of sauce per can. They were then able to calculate that in a days production on this line the sauce loss was $106.00.

To prevent spoilage in canned onions, it is essential that the pH be maintained below 4.5. This is done by the addition of citric acid to the brine. Lot Plot Fill Evaluation studies were made to determine the Upper Lot Limit weight of fill. Sample cans were filled with this weight of onions and increasing amounts of citric acid added to the brine. The citric acid brine which adjusted the sample can to a pH of 4.45 was used in canning the onions.

USING THE LOT PLOT METHOD

1. The precision of the scale or balance to be used in the test work must be established as well as the cell width (reading interval). This is done by measuring the net or drained weights of the first five of the 50 cans sampled. The highest and the lowest weights are recorded as well as the difference between them. This weight difference should be multiplied by 2 and divided by any number between 7 and 14 that will give a readable scale division. This division or weight interval, is designated as a cell. Where weights fall in between weight divisions, they should be recorded at the lower weight limit in every case. After all 50 weights are recorded on the Lot Plot form they should spread vertically over no fewer than 7 cells nor more than 14 cells. For example, in Figure #2 the maximum difference of the first five weights (each designated as #1) was .20 ozs. (between 6.50 and 6.70 ozs.); doubling this value gave .40 ozs. This figure was divided, in this case, by 8 and a cell width of 0.05 ozs. was obtained. The scale used for weighing these samples had an accuracy, perforce of at least 0.05 ozs.

2. In the estimated Cell Number column on the form, the cells are numbered from 1 to 10 above the zero point and from -1 to -10 below the zero point. When using, insert the weight interval readings in the value column on the Lot Plot form, placing them so that the average of the first five cans (in this case 6.60 oz.) is opposite zero point in the Cell No. Column. Enter the first five samples weighed at the proper point on the chart designating them as "1". The second five samples weighed, designate as "2". This is continued until all of the 10 sets of 5 weights each are recorded on the form. For convenience, the Roman numeral X is used in place of 10.

3. A rapid method to find the average of the 50 samples weighed is as follows: Refer now to the Lot Plot Fill Evaluation Report (Fig. 2).
A. In a normal distribution, the largest number of weights will be found in the zero Cell Row. If this is true on the chart, proceed to step B. If this is not true, cross out the number of cell values and renumber the cells in the final Cell column of the form, so that the maximum number are in this position.
B. Mark a zero in the Calc. Ave. Column on the form in the zero Cell Row. Compare the number of samples found in the plus one cell and minus one cell row, and show the numerical difference in the calculated

Figure 2

LOT PLOT FILL EVALUATION REPORT

Date March 8, 1954
Plant #1
Line No. #2
Filler A-10-P
Filler Speed C.P.M. 255

Can Size 300 x 407
Product Pork + Beans
Remarks

Drained Weight x Net Weight

Value	Final Cell No.	Est. Cell No.	1	2	3	4	5	6	7	8	9	10	20 Limits & Spec.	Calc. Ave.
oz.														
7.10		10												
7.05		9												
7.00		8												
6.95		7											U.L.L.	
6.90		6												
6.85	4	5	2	7	9									+8
6.80	3	4	5	6	7	8								–
6.75	2	3	3	4	4	5	6	7	8					+4
6.70	1	2	1	2	3	3	4	8	8	9				–
6.65	0	1	2	5	5	6	6	7	8	9	X	X	$\overline{X}$	0
6.60	-1	0	1	3	3	5	6	9	X	X				–
6.55	-2	-1	1	4	7	9	X							–
6.50	-3	-2	1	1	2	4								–
6.45	-4	-3	2											–
6.40		-4											L.L.L.	
6.35		-5												
6.30		-6												
		-7												
		-8												
		-9												
		-10												
													ΣX	+12
													$\overline{X}$	+.24

Range	
1	4
2	8
3	3
4	5
5	4
6	4
7	6
8	3
9	6
X	2
ΣR	45
3σ	5.8

Center Zero Cell	6.675
Cell Interval x $\overline{X}$ .05 x (+.24) =	+.012
Average Fill	6.687 oz.
Cell Interval x 3σ .05 x 5.8 =	.290
Upper 3σ Lot Limit	6.977 oz.
Lower 3σ Lot Limit	6.397 oz.
Calc. Range of Fill (6σ)	.580 oz.

Strays

[redacted]	Number Found
Low	None
High	None

Remarks:

Signature L.G + F. J.

TABLE I

Table for converting sum of Range (ΣR) to 3σ for sample size 50.
Sum of Range of 10 sub-groups of sample size 5.

Sum ΣR	3σ Cells	Sum ΣR	3σ Cells	Sum ΣR	3σ Cells	Sum ΣR	3σ Cells
15	1.9	31	4.0	47	6.1	63	8.1
16	2.1	32	4.1	48	6.2	64	8.3
17	2.2	33	4.3	49	6.3	65	8.4
18	2.3	34	4.4	50	6.5	66	8.5
19	2.5	35	4.5	51	6.6	67	8.6
20	2.6	36	4.6	52	6.7	68	8.8
21	2.7	37	4.8	53	6.8	69	8.9
22	2.8	38	4.9	54	7.0	70	9.0
23	3.0	39	5.0	55	7.1	71	9.2
24	3.1	40	5.2	56	7.2	72	9.3
25	3.2	41	5.3	57	7.4	73	9.4
26	3.4	42	5.4	58	7.5	74	9.5
27	3.5	43	5.5	59	7.6	75	9.7
28	3.6	44	5.7	60	7.7	76	9.8
29	3.7	45	5.8	61	7.9	77	9.9
30	3.9	46	5.9	62	8.0	78	10.1

Calculated from $3\sigma = 3\bar{R}/d_2 = (3/2.326)\bar{R} = (1.29/10)\Sigma R$

---

average column. Place a plus or minus value accordingly. Multiply the difference between the number of samples found in the plus 2 cell and the minus 2 cell row by two and record. Proceed similarly through the remaining cells, multiplying the difference by the cell number. For example, in figure 2 eight samples were found in both the plus one cell and minus one cell rows. Under these conditions, there is no need for a notation in the Calc. Av. column. In the plus 2 cell row, there are seven samples and in the minus 2 cell row, 5 samples giving a difference of plus 2. This is multiplied by 2 and entered.

C. Add the plus values in the Calc. Av. column and subtract the minus values and record the sum at point ΣX being sure to denote the proper sign (+ or -).

D. Divide the ΣX value by 50 and record at $\bar{X}$.

E. Determine the weight value at the center of the Final Zero Cell and record it. In as much as all weights are recorded to the next lowest weight interval, the center of the Final zero cell will lie half way between its indicated weight, and that of the plus one cell.

F. Enter the Cell Interval (.05) multiplied by $\bar{X}$ value (+.24) on the form.

G. Subtract or add the value found in (F) above (depending on the sign) to obtain the Average Fill weight of the 50 samples.

H. Locate the average fill weight on the form and mark $\bar{X}$ in the Limits and Specification column.

4. Calculate the Range of the Fill (actual spread of weights) as follows:

A. Observe the 50 recorded weights on the form. If the recorded weights give a reasonably symetrical bell type distribution (Figure 1), proceed to step B. The sets of weights (numbered 1 to X) should have a random distribution if the process is statistically stable during the test period. If the distribution is not reasonably normal refer to methods described by Dorian Shainin (7,8).

B. Observe the first five weights which are marked "1" on the form. Count the number of cells vertically from the lowest cell which "1" occupies to the highest cell occupied by "1", not including the lowest

cell and record opposite 1 in the Range column. Similarly find and record the range for the sets of 5 weights between 2 and 10.
C. Add the 10 Range values and enter this total after ΣR.
D. Convert ΣR to 3σ by referring to Table 1 and record.
E. Enter Cell Interval (.05) X 3σ (5.8) on the form.
F. Subtract the above value from the Average Fill to obtain the Lower Lot Limit. Add the same figure to the Average Fill to obtain the Upper Lot Limit.
G. The Range of Fill is obtained by subtracting the Lower Lot Limit from the Upper Lot Limit.
H. Enter the Lower Lot Limit (L.L.L.) and the Upper Lot Limit (U.L.L.) on the form in the Limits and Specifications column.

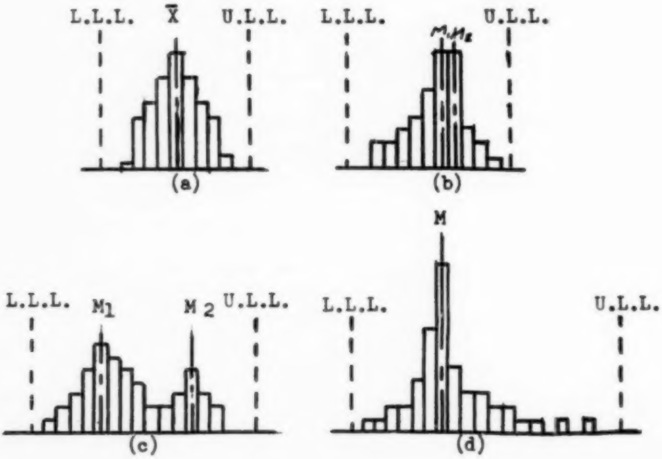
METHODS OF CALCULATING RANGE OF FILL WEIGHTS AND LOT LIMITS

After a Lot Plot form has been completed, it may be turned ninety degrees for observation. In this position it takes the form of a distribution chart.

1. Where a normal bell shaped distribution is found, as shown in figure 3 (a), proceed as previously indicated in Using the Lot Plot Method, paragraph 4 above.

2. The Range of fill and the Lot Limits of non-symetrical distribution as shown in figure 3 (b,c & d) may be estimated by the Half Distribution method described by Dorian Shainin (7,8). When non-symetrical distributions are found, an investigation should be made to determine the reason for their existence. For example, one or more of the filler pockets may be out of adjustment which may cause wide fluctations in the fill weights. When this condition is corrected, the Range in the fill weights will be reduced, and the distribution curve will be of a normal bell shape.

Figure 3



STRAYS

A stray occurs when an occasional weight falls outside of the normal expected pattern or Lot Limit. Strays are difficult to handle in statistical calculation. They are definitely important in Fill Evaluation Studies and every effort should be made to determine why they are occurring. In one test where strays were encountered it was found that the filler was slightly out of time with the closing machine. This caused occasional spillage and subsequent stray readings.

TESTS FOR SIGNIFICANT DIFFERENCE

When comparing two Lot Plot Fill Evaluation reports, visual observations of the frequency distributions will, in many cases, be sufficient. If the differences are small, for a normal distribution, the following tests may be used. For additional tests or for different ratios when the Lot sizes differ from 50, refer to Dr. Duncan's (9) book on Quality Control and Industrial Statistics.

1. Test for significant difference in standard deviation when comparing two Lot Plot Fill Evaluation Reports.

Level of Significance (α)	F (one tail test)	F (two tail test)
0.1	1.4	1.6
.05	1.6	1.8

$$\frac{(\sigma_1 \text{ largest})^2}{(\sigma_2 \text{ smallest})^2} =$$ This value must be larger than F values shown to be significant.

Use F one tail test value when testing for improvement.
Use F two tail value for general testing.

2. Test for significant difference in average fill weights when comparing two Lot Plot Fill Evaluation Reports.

Level of Significance (α)	Critical Value (one tail test)	Critical Value (two tail test)
0.1	+1.28	+1.65
.05	+1.65	$\pm$1.96

$$\frac{\bar{X}1 - \bar{X}_2}{.141\sqrt{\sigma_1^2 + \sigma_2^2}} =$$ must fall outside critical values given above to be significant at levels shown.

Use critical value (one tail test) when testing for improvement.
Use critical value (two tail test) for general testing.

EXAMPLE: The following results were obtained from Lot Plot Fill Evaluation Reports on 303x406 Lima Beans.

Filler Speed C.P.M.	Average Fill Weight ($\bar{X}$)	Calc. Range of Fill (6σ)	Standard Deviation (σ)
130	12.60 oz.	5.0 oz.	.833
170	12.38 oz.	5.6 oz.	.933

$$\frac{(\sigma_1 \text{ largest})^2}{(\sigma_2 \text{ smallest})^2} = \frac{.933^2}{.833^2} = 1.25$$

No significant difference at the 0.1 level was found in the standard deviation.

$$\frac{\bar{X}_1 - \bar{X}_2}{.141\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{12.60 - 12.38}{.141\sqrt{.833^2 + .933^2}} = 1.29$$

A significant difference at the 0.1 level was found in the average fill weights. A one tail test is applicable in this case since it is known that fill weights usually decrease as the filler speed is increased.

SETTING SPECIFICATIONS AND LINE CONTROL PROCEDURES FOR FILLING OPERATIONS

The information obtained from the foregoing work may be used to design control procedures for filling operations. A single total weighing of an established number of units will suffice if the information obtained from the Lot Plot Form is utilized. A method to obtain Specification limits and set up a line control procedure is as follows:

1. Determination of Sample Sizes.

It is necessary to pre-determine the number of cans that must be used to establish whether an adjustment in the fill is necessary.

2. Determination of the Aimed at Average ($\bar{\bar{X}}'$).

A decision must me made as to where to place the Aimed at Average ($\bar{\bar{X}}'$) in relation to a Given Specification Limit. Referring to point (a) figure 4, if the specification is set at that point, 99+% of the cans filled will weigh more than the required minimum. If set at point (b) 98% of the cans filled will weigh more than the minimum Given Specification Limit. If set at point (c) 50% of all cans filled will weigh less and 50% weigh more than the Given Specification Limit.

3. Calculation of Control Limits.

The Control Limits may be calculated, using the following formulae. The Control Limits shown below are two standard Deviation Limits of the average. They apply in cases where fillers can easily be adjusted. If it requires considerable time to make filler adjustment, substitute 3 in place of 2 in the formulae.

Lower Control Limit $= N(\bar{\bar{X}}' - 2\sigma/\sqrt{N})$

Upper Control Limit $= N(\bar{\bar{X}}' + 2\sigma/\sqrt{N})$

The sample size N is determined as described in paragraph #1.

The Aimed at Average ($\bar{\bar{X}}'$) is determined as described in paragraph 2.

The Standard Deviation is determined from Lot Plot Fill Evaluation Reports.

The above formulae apply only to a normal distribution.

SUMMARY

The Lot Plot method of Fill Evaluation is a convenient tool for those interested in studying or controlling filling operations. Here the main interest is in two specific things, (1) The Average Fill Weight and (2) The distribution of the individual fill weights, especially the lowest and highest. The difference between the highest and lowest fill weights is called the range.

Where only fifty cans are used for check weighing, it is probable that neither the lowest or highest fill weight which is being shipped out to customers will be found. In most cases, however, these can be rather accurately calculated by Lot Plot Fill Evaluation Studies. The latter may be readily used to test the efficiency of two different types of fillers. They may also be used to test mechanical improvements or different operation conditions to see if better filling may be attained. The information obtained from such studies may be used as a basis for setting up rapid and simple line control procedures.

Literature Cited

1. E. McKinly   Nat. Canners Assoc. News Letters 1472 p.69 Jan30,54
2. I. MacPhail   "   "   "   "   "   " p.70 " " "
3. H. Link & H. Dobson"   "   "   "   "   " p.71 " " "
4. C. B. Way   "   "   "   "   "   " p.73 " " "
5. H. I. Edwards   "   "   "   "   "   " p.76 " " "
6. W. A. Brittin   "   "   "   "   "   " p.77 " " "
7. Dorian Shainin Ind. Quality Control p. 15 Vol VII, No. 1, July 1950
8. "   "   Transactions Middle Atlantic Regional Conference
   Am. Soc. Quality Control, p. 31, Feb. 5,6, 1954
9. A. J. Duncan Chapt. 20, Quality Control Ind. Statistics

RATING SCALES AND PSYCHOLOGICAL FACTORS
IN TASTE PREFERENCE RESEARCH

James A. Bayton
Howard University

The problem of quality control of the taste of food products is a coin with two sides. On one side there is the attempt to control taste quality in terms of already determined standards. The other side of the coin is the determination of the standards in the first place. This paper deals with the latter aspect of the problem. Our premise is that preference data based upon representative samples of the consumer population should be among the elements contributing to standards for food quality.

The methods for ascertaining taste preferences fall into two basic psychological categories--the Method of Comparative Judgments and the Method of Single Stimuli. With the Method of Comparative Judgments the various items being evaluated are presented to the subjects within one session and direct comparisons are made between the items. In contrast, with the Method of Single Stimuli each item is judged in "isolation" without any specific comparison stimulus being present (1,3).

One of the limiting factors in the use of comparative judgment procedures in taste testing with non-expert subjects is the relatively rapid adaptation rate for taste. The adaptation factor restricts the number of tastings per session for each person. It has been our experience that with a paired comparisons design only three items can be tasted per session since three pairings (six tastings) are involved. With four items a paired comparisons design calls for six pairings (12 tastings). An alternative comparative judgment method would be a rank order design. However, we doubt that more than four items can be evaluated per session by a rank order design without risking difficulties from adaptation.

There is an even more important consideration that enters the picture when one is trying to determine consumer preferences as a factor in quality control of the taste of food products. Which of the two approaches--Method of Comparative Judgments or Method of Single Stimuli--most nearly approximates the typical situation of the consumer? It seems reasonable to contend that seldom does the consumer have available at a given moment in time several variations of a food product for direct comparative evaluation. The consumer usually has only one of the possible variations at a given moment and time. He tastes the item and judges it against the general background of his accumulated experience. This circumstance is a duplication of the Method of Single Stimuli. It follows, therefore, that naturalistic or realistic research on consumer taste preferences demands that a Method of Single Stimuli approach should be employed rather than comparative judgment methods such as paired comparisons and rank ordering.

When the subjects are to judge only one item per session some type of rating scale must be provided. There are different kinds of rating scales, however, and one is faced with the problem of selecting the most efficient for use under real-life, home conditions in a consumer survey. The following experiment was designed to evaluate three rating scales (3). The products used in the test were three canned orange juices that varied in Brix-acid ratio with degrees Brix constant. The

three juices ranged from tart through sweet. One of the scales was the following 7 point scale (the scoring is shown in the parentheses):

(7) Excellent	the best canned orange juice I have ever tasted
(6) Good	much better than other canned orange juices I have tasted, but not the best
(5) Fair	a little better than other canned orange juices I have tasted, but not much better
(4) Borderline	can't decide whether it is better or worse than other canned orange juices I have tasted
(3) Poor	a little worse than other canned orange juices I have tasted, but not much worse
(2) Very poor	much worse than other canned orange juices I have tasted, but not the worst
(1) Objectionable	the worst canned orange juice I have ever tasted

The second rating scale was of the "thermometer" type (Fig.1). The subjects were instructed to decide first what they thought of a juice in a general way--"Very Good," "Poor," etc.--and then to rate it by assigning a score in the particular area selected. The third scale was an adaptation of a scaling procedure that has been used with success in opinion research in social psychology. We call this an unstructured scale (Fig.2). Only the ends of the continuum are defined; the subjects were shown that their reaction to a juice could be expressed as falling anywhere from "Very Poor" up through "Excellent." A cross was to be put in the square that expressed opinion about the juice.

The experiment was conducted in a panel of 90 households randomly divided into three sets of 30 households each. Each set of households worked with only one scale and evaluated only one juice per session. Several days intervened between placements of the three juices. The order in which the three juices were placed varied randomly throughout the panel. After one and two months intervals the panel members were retested in order to investigate the reproducability of the original data. The results of this experiment indicated that the unstructured scale (Fig.2) was the most efficient in terms of the statistical significance associated with the preference patterns obtained and in the reproducability data.

On the basis of these results the unstructured scale was used in a study of preferences for six canned orange juices that varied in Brix-acid ratio (2). The sample was 720 randomly selected households in Indianapolis. The juices tested were 12, 14, 16, 18, 20, and 22 Brix-acid ratio. The first week each household received one of the six juices. One week later each household was given another juice. These assignments were made in such a manner that every possible combination of two juices occurred in the two test sessions. Without being informed, some households received the same juice for the two test sessions. Each person in a household who was 16 years of age and over rated a juice on each of three days. The data were analyzed in terms of the three-day means per individual.

When the mean preference ratings for the respective juices were obtained no sharply defined pattern of preference was observed, in spite of the fact that these juices varied from rather tart to very sweet.

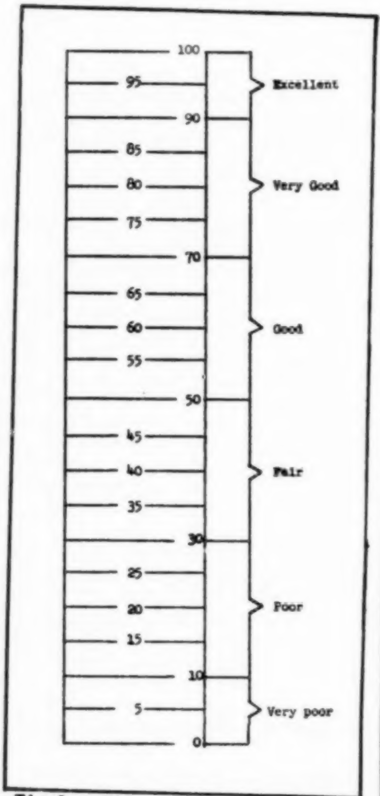


Fig.1 - Thermometer-type Scale

Fig.2 - Unstructured Scale

However, prior experience had shown that a more intensive analysis was required in the search for preference patterns. Accordingly, all persons who had judged a given juice were divided into two groups--those who scored it above the mean for all subjects and those who scored it below that mean. The former were called the "Like" group, the latter a "Dislike" group. Once these groups were isolated the ratings given to the combinations of paired juices were inspected and this is the level at which preference patterns emerged.

The first pattern noted was that those who "liked" any one juice also "liked" the other juices, whether they were tart or sweet. Those who "disliked" any one juice also "disliked" the other juices. We concluded, therefore, that with respect to canned orange juice there are two general groups of consumers--those favorably disposed to this product and those not so favorably disposed, regardless of the tart-sweet characteristics of the various juices.

Within these general "liking" and "disliking" groups, however, there were what might be termed second-order preference patterns. Furthermore, the second-order patterns were different in the "like" and "dislike" groups. Within the general "like" group the same individuals showed equal preference for two different juices--one relatively tart and the other somewhat sweet. This is seen in Fig.3 which shows how those who "liked" 12 Brix-acid ratio rated the other juices. From 12 through 16 Brix-acid ratio the preference ratings decreased; at 18 Brix-acid ratio the ratings became relatively high again. From that point on the preference ratings declined once more. In terms of a significance test based on the means of the individual differences, 12 and 18 Brix-acid ratio were not different in degree of preference. Apparently, consumers who are favorably disposed to canned orange juices in general expect them to be either somewhat tart or somewhat sweet and within each region there is a preferred juice.

The second-order preference pattern within the general "dislike" group indicated that this group was really composed of two different sub-groups. One of the sub-groups showed highest preference only for a tart juice; the other sub-group preferred only a sweet juice. (Note the contrast to the general "like" group in which the same people preferred two different juices). In Fig. 3 we see that among those who "disliked" 12 Brix-acid ratio the preference ratings increased up to 18 Brix-acid ratio and then decreased. Fig. 4 shows that those who "disliked" 20 Brix-acid ratio exhibited a tendency to increase their preference ratings as the juices became more tart. If the general "dislike" group is in fact composed of two different sub-groups--one preferring a tart juice and the other preferring a sweet juice--an analysis based upon a juice in the center of the series should reveal A U-shaped preference function. That such is the case is seen in Fig. 5. Among those who "disliked" 18 Brix-acid ratio preference increased when the paired juice was more tart; preference also increased when the paired juice was sweeter.

This research led to the following recommendation to the citrus industry. Two different kinds of canned orange juice might well be marketed--one relatively tart at 12 Brix-acid ratio and one relatively sweet at 18 Brix-acid ratio. People already favorably disposed to canned orange juice would find both of these juices acceptable. The tart juice would be available for those who at present are not so favor-

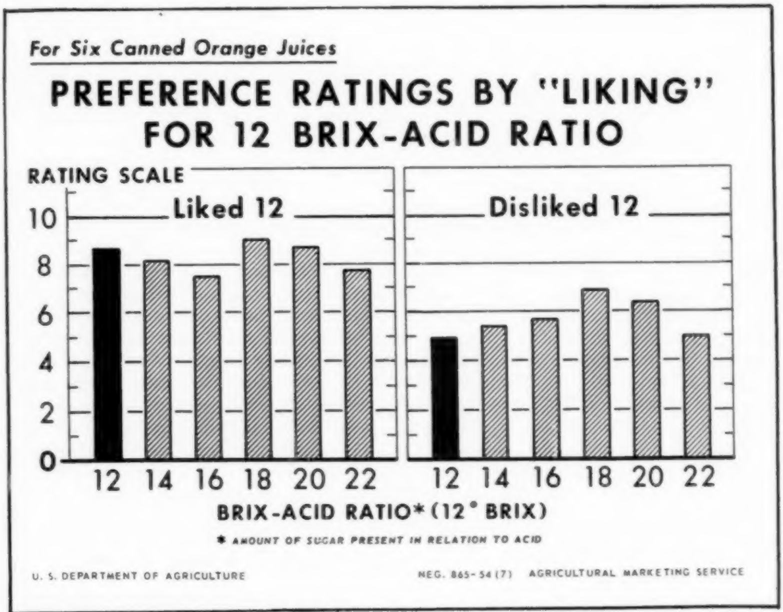


Fig. 3 (U.S.D.A. Photograph)

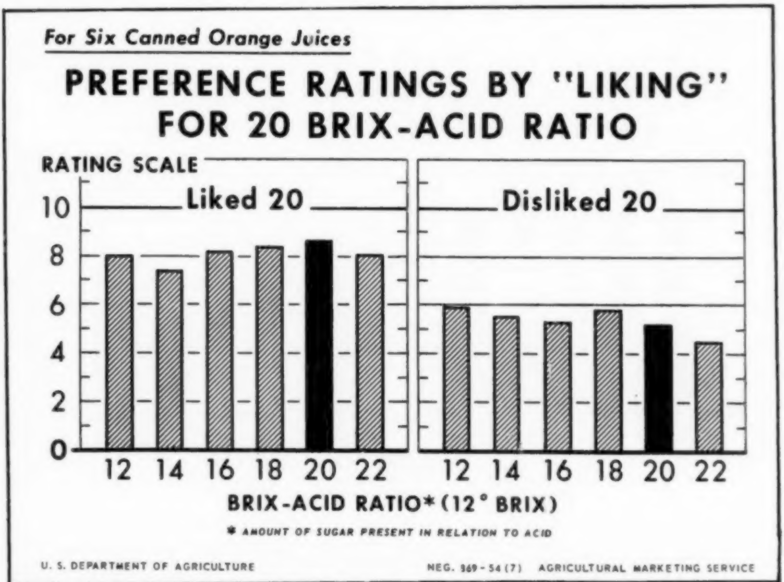


Fig. 4 (U.S.D.A. Photograph)

ably disposed to these juices but who do prefer a tart juice exclusively. A similar circumstance would exist with respect to the sweet juice.

We have two types of evidence as to the validity of the unstructured scale. As stated the canned orange juices varied from tart through sweet. The subjects were not informed that this was the variable under investigation. At the end of each session with a given juice the subjects were asked to check those items on a list which they thought were most descriptive of the juice. Favorable comments about a juice--"just the right tartness," "just the right sweetness," etc.--always yielded higher percentages for the "like" groups than the "dislike" groups. The validity of this scale is seen also in its correlation with the answers to this question which was asked after each juice was tested: "If a juice that tastes like this one was on the market, would you like to have it served here in your home?" For those who scored 12 Brix-acid ratio above the mean for all subjects rating that juice, 83 percent said, "Yes." Among those who scored this juice below the general mean, 21 percent said, "Yes." The answers to this question correlated perfectly with the preference patterns for the "like"--"dislike" analysis. Whereas 83 percent of those who "liked" 12 Brix-acid ratio answered in the affirmative, among those who had 12 and 16 Brix-acid ratio 53 percent said, "Yes," for the latter juice. For the 12--18 Brix-acid ratio combination, 94 percent gave an affirmative answer for the 18 Brix-acid ratio juice. Within the "dislike" group the percentages of affirmative answers followed the pattern revealed by the rating scale.

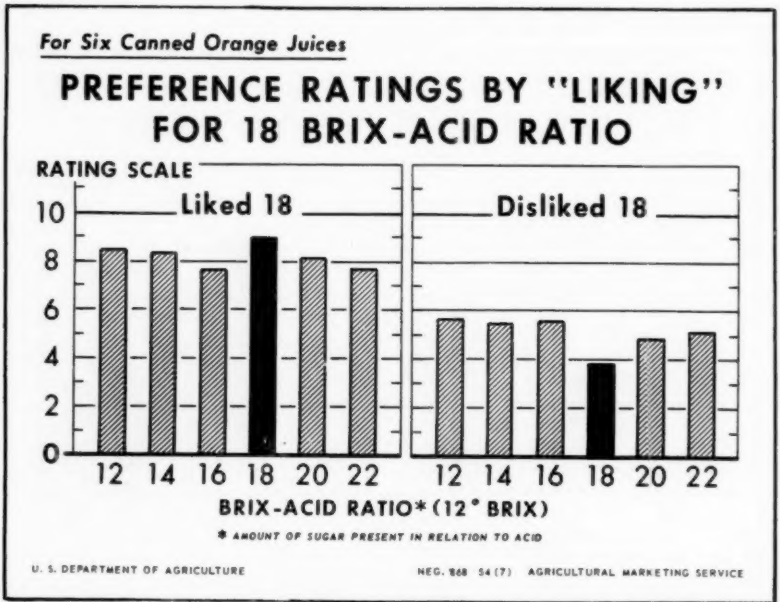


Fig. 5 (U.S.D.A. photograph)

The data from the earlier 90 household experiment and from the 720 household project showed that this scale has a high degree of reliability with respect to preference patterns. There is some evidence, however, that a frame of reference factor can operate in some instances to change the general level upon which the products are being judged. In one experiment with a Method of Single Stimuli design it was found that when subjects had prior experience with the juices they tended to assign higher ratings. However, the preference pattern between juices was not affected (1). In an unpublished project on preferences for white pan breads that varied in specific volume, milk solids, lard, and sucrose it was found also that as the subjects worked from week to week, but one bread at a time, the general level of the scoring tended to change. That this does not always occur was seen in the Indianapolis study. Those subjects who had the same juice in the two sessions tended to give substantially the same rating each time (2). The existence of a frame of reference factor cannot be detected with Method of Comparative Judgment designs such as paired comparisons and rank order. In the latter cases the judgments are such that direct comparisons between items is all that is obtained. Whether the entire set of items can shift to a more favorable or to a more unfavorable position is not known.

A legitimate question is whether the preference patterns obtained with the Method of Single Stimuli procedure are similar to, or differ from, those obtained with Method of Comparative Judgment designs. Data on this point indicate that the two general procedures do yield different preference patterns. In one experiment with canned orange juices that varied both in degrees Brix and in Brix-acid ratio a rank order design produced preference differences in terms of degrees Brix and Brix-acid ratio. A Method of Single Stimuli design produced preference differences only in terms of Brix-acid ratio (1). Morse (4) has studied preferences for the same six canned orange juices used in the Indianapolis project. He used the same unstructured scale but later had his subjects rank the juices in order of preference. Morse reports that with the scale his results were similar to ours in that the mean ratings for all subjects per juice were not different from 12 through 20 Brix-acid ratio. The 22 Brix-acid ratio juice was given a lower rating. On the other hand, the rank order procedure yielded a curvilinear function with highest preference at 20 Brix-acid ratio and low preference at 12 and at 22 Brix-acid ratio.

What can we say about such different results? First of all, it is our contention that the Method of Single Stimuli results are more valid because of the realism of the testing situation--one item is judged per session. Secondly, we believe that the analysis in terms of "liking" and "disliking" categories gets at hidden aspects of the preferences of consumers for canned orange juices that would not be exposed readily by the rank order design.

Another problem is encountered when rating scale data obtained under Method of Single Stimuli conditions are used for inferences about the discrimination function in taste testing. It is a generally accepted principle in taste preference research that items should be used which are readily distinguishable for the subjects. The determination of discriminable items is usually done with duo-trio and triangle tests. Working with canned orange juices and non-expert subjects we have found consistently that a Brix-acid ratio difference of four is necessary for discrimination with a duo-trio test. Note, of course, that both duo-trio and triangle tests are within the Method of Comparative Judgment

category. In the Indianapolis project--using the unstructured scale under Single Stimuli conditions--the data indicate ability to discriminate at only 2 Brix-acid ratio difference between juices. For example, in the "like" 16 Brix-acid ratio group which had 18 Brix-acid ratio as the paired juice the mean difference for the preference ratings of the two juices was statistically significant. One is forced to conclude that these two juices, only 2 Brix-acid ratio apart, were discriminable for the particular subjects involved. There was a suggestion, however, that the ability to discriminate between the juices was greater among the various "like" groups--those favorably disposed to all of the six juices--than among the various "dislike" groups. Of the ten instances in which the pair of juices rated was only 2 Brix-acid ratio different, there were nine cases, for the "like" groups, that produced significant differences in preference ratings. In contrast, this was true for only five of the ten instances among the "dislike" groups. The project dealing with consumer preferences for white pan breads also gave results suggesting that finer discrimination occurs with the unstructured scale in a Method of Single Stimuli procedure than is true under a duo-trio design.

The use of a rating scale to infer discrimination could, however, be misleading in some instances. It will be recalled that in the "like" 12 Brix-acid ratio group those who had 18 Brix-acid ratio as the alternate juice gave the two juices ratings which were not significantly different. On the surface this seems to indicate that the subjects could not distinguish between the two juices. Actually, all evidence shows that these two juices are easily distinguished--it just so happens that they are equally preferred in spite of the difference in taste.

In summary, the following points can be made: 1. Consumer taste preferences should be a factor in the standards used in quality control. 2. The Method of Single Stimuli approach to determination of consumer taste preferences is more realistic than any of the various Method of Comparative Judgment approaches. 3. An unstructured scale, with only the ends of the continuum defined, is a valid and reliable tool for ascertaining consumer taste preferences with a Method of Single Stimuli design.

References

1. Bayton, J. A. and Thomas, C.M. Comparative and single stimulus methods in determining taste preferences. Journal of Applied Psychology, 1954, 38, 443-445.

2. Bayton, J. A. and Bell, H.P. Preferences for canned orange juices that vary in Brix-acid ratio. Agricultural Marketing Service, Market Research Report No. 76, 1954 (United States Department of Agriculture.)

3. Clements, F.E., Bayton, J.A., and Bell, H.P. Method of single stimulus determinations of taste preferences. Journal of Applied Psychology 1954, 38, 446-451.

4. Morse, R. L. (Personal Communication.)

FLEXIBLE BUDGETS - THE SOUNDEST WAY OF CONTROLLING Q.C. COSTS

Richard H. Stewart
Lear, Incorporated

Budgets aren't usually considered a very palatable subject by most people, whether in Q.C. or any other field of industrial endeavor. They are generally felt to be as dry as the Mohave Desert, irritating as a cinder in your eye, and as clear as Einstein's Theory of Relativity. I won't argue the point since I shared the same opinion not too long ago. However, I think we at Lear have arrived at a means of controlling Q.C. costs which is the least objectionable, the most easily understood, and the most fruitful in point of obtainable results. My purpose in the following discussion will be to highlight an evaluation of our budget control and leave you with a few ideas which will assist you in your own budget problems.

Our early efforts to control costs were geared to a system which used a projected sales forecast as a starting point. Our Contracts Division developed a schedule for the coming year in terms of:

1. Signed contracts on the books.

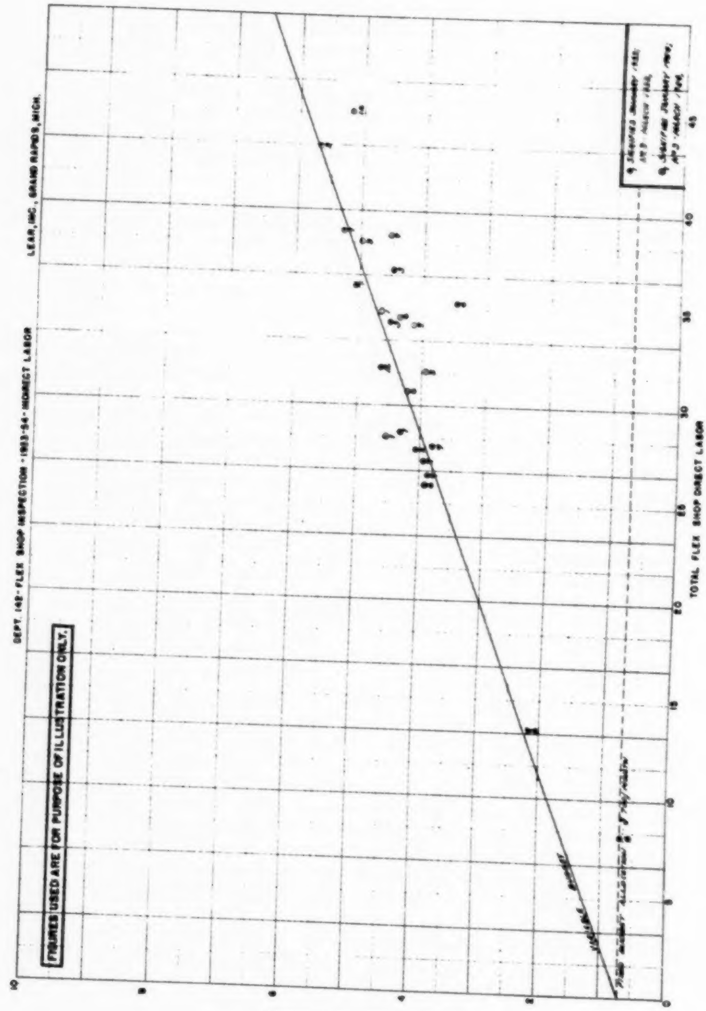
2. Contracts not yet formalized but carrying a high degree of assurance of finalization within the forecast year.

This information was first transmitted to Production management who established their Direct Labor needs. Q.C. then took the Direct Labor estimate and using previously developed manpower ratios arrived at its personnel needs. To this figure was added the requirements for so-called "non-production inspection" (including tooling, receiving inspection, clerical, Q.C. analysis and administration). The total constituted our best appraisal of an overall Q.C. forecast. This figure, as amended by the plant manager, then became the Q.C. budget.

At this juncture it should be emphasized that the budget did <u>not</u> provide for changes from the forecast, either up or down. In other words, there was no direct relation between the budget and what could be termed an "activity or workload index". To our way of thinking the lack of such a relationship was the biggest shortcoming in our original cost control efforts, since a change in workload did not result in a change in budget.

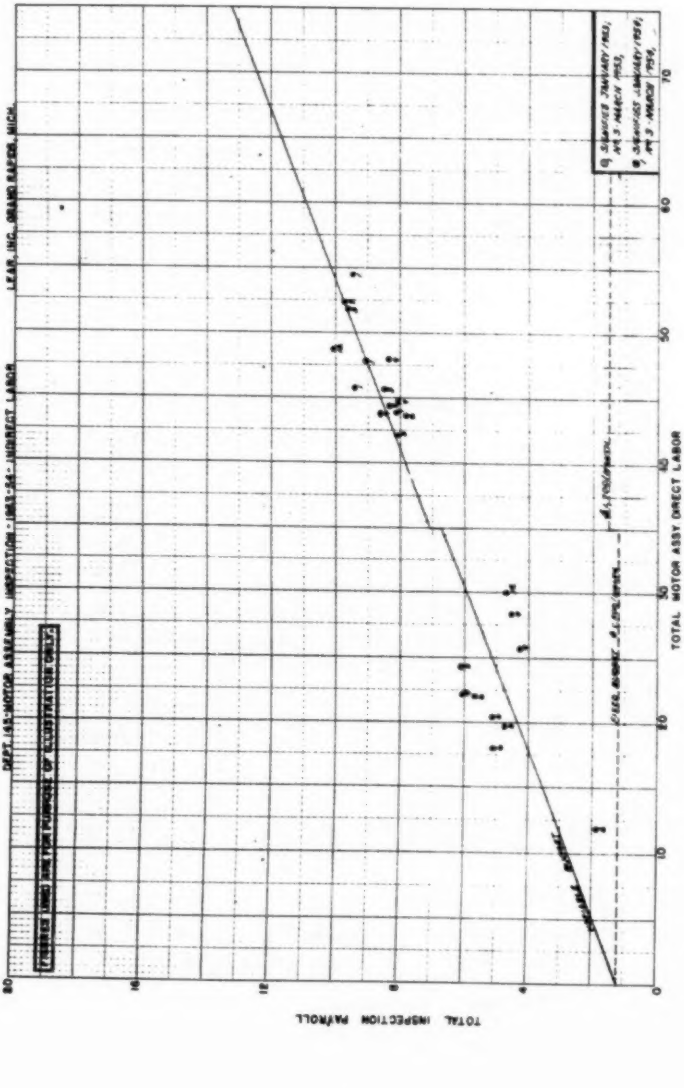
Realizing this shortcoming, we set about to improve the procedures. Our first move was to take each major cost area, namely, Machine Shop, Electronic Assembly, Gyro Labs, Motor Assembly, etc., and develop a "scatter diagram". FIGURE I on the following page illustrates the procedure used. The chart is graduated vertically for "Inspection Payroll" and horizontally for "Direct Labor". In Production Departments the Direct Labor payroll is considered the most accurate, currently accessible figure which was a reasonable rule of the inspection workload. By picking points for each month during the review period, a pattern is established to show what our performance has been under a given set of conditions. By points I mean Inspection costs as a function of Direct Labor costs. A line is then drawn from the lower left hand area through the points forming the expenditure grouping. You'll note the starting point is not the inter-section of the ordinate and abscissa............

FIGURE I



LEAR, INC., GRAND RAPIDS, MICH.
FIGURES USED ARE FOR PURPOSE OF ILLUSTRATION ONLY.
TOTAL FLEX SHOP DIRECT LABOR
TOTAL INSPECTION PAYROLL

FIGURE II



LEAR, INC. GRAND RAPIDS, MICH.
TOTAL MOTOR ASSY DIRECT LABOR
TOTAL INSPECTION PAYROLL

I'll explain why later. The line thus drawn is not a weighted average but rather a compromise between the highest and lowest points in the group. Accounting Budget Department and the Inspection Departments try to set the budget line at a level that experience shows is an attainable, though difficult, payroll cost level. The starting point on the abscissa is above the inter-section of the ordinate and abscissa because at the lowest level of Direct Labor expenditures in this particular department the minimum inspection costs have been estimated at $700 per month.

FIGURE II is the same as FIGURE I, except that it contains an increase of inspection costs at Direct Labor of $35,000. The change is caused by an increase in the fixed portion (inspection supervision) required at this level of Direct Labor. A procedure similar to that just explained is followed in each of the other cost areas.

The next step in the program is to provide Department Heads with the information they need to plan their work. Instead of giving them a copy of each master chart, we furnish a corresponding manhour budget. FIGURE III is an example of the card which records manhours provided, budget variance, premium hours paid and roster strength (number of people receiving checks). The manhours allocated include all manhours

FIGURE III

QUALITY CONTROL AND INSPECTION DIVISION

PAYROLL – HOURS REPORT Dept. No. 142

PERIOD	ACTUAL HRS. WORKED	BUDGET VARIANCE	PREMIUM HOURS PAID	BUDGET VARIANCE	ROSTER STRENGTH
1-23-55 Week Ending	503.8	(51.2)	57.4	(17.4)	7 days
Month to Date	1654.4	-134.6	212.1	-43.9	3 nites

Reasons for variance: Worked 4 inspectors Saturday, 1-22-55, to complete rush order No. 56789. This order required on assembly line Monday, 1-24-55.

Figures Given Are For Illustration Only H Valenwood
ASS'T. CHIEF INSPECTOR 1-24

LEAR 50.1-1

actually worked, regardless of whether they are worked at a straight time or premium (overtime) rate. The following hours are excluded:

1. Holiday pay.
2. Premium portion of overtime (whether at time and one-half or double time).
3. Absenteeism.
4. Leaves of absence.

Each card covers a specific department for a particular week, and also records the expenditures for the month to date. Provisions are also made for weekly expenditures and variances from budget. Department Heads prepare Payroll-Hour Reports each week covering their areas of responsibility. These are turned in to the Division Manager for review.

As a further aid in the cost control program, a Variable Overhead Budget Performance Report is issued each Thursday by the Budget Department of the Accounting Division. It covers the performance of the week ending the preceding Sunday (FIGURE IV). Several advantages accrue from such a report. First, it furnishes a performance record in dollars. Second, it relates these costs to corresponding Direct Labor. Third, it indicates whether we have over or under spent, and the amounts. Fourth, controllable and uncontrollable expenses are also included. Additional points about the Variable Overhead Budget Performance Report also bear discussion. You will note that under Indirect Payroll - Variable, the Budget allowance is $1,420, the permitted 12.7% of the Actual - Department Direct Labor Payroll - and not 12.7% of the budgeted figure. Also to be noted is the figure of $167 for Budgeted, which is the same as that for Actual. The reason for this is that any charges over the $167 in the Fixed Actual are added to the Variable figure under Actual. Vacation pay is a prorated figure, which in some cases may be an out of period expense but is nevertheless reported.

Receiving and Tool Inspection Budget provisions presented a slightly different set of conditions than those incurred in the aforementioned areas. This was because inspection costs in these areas could not be related to the same yardstick as that used in Production Departments. Therefore, the Receiving Inspection budget was formulated in the following manner.

The material receipts for the past twelve (12) months were first established from available records. These included material from within the plant requiring processing by Receiving Inspection, (heat treat checks, identification, magnaflux, certain gear rolling, and electrical checks), in addition to that obtained from sub-contractors.

Dividing the number of lots received by the total manhours spent clearing them gave us the hours per lot. Then by working backwards and using the forecast of material receipts, we were able to estimate our manpower needs in much the same manner we did with the production forecast. These total requirements consisted of two basic elements.....one fixed and the other variable. Supervision, clerical, parts handlers, rejected material handlers, certified test report coordinators and sorting comprised the fixed portion, and direct inspection at straight time the variable. FIGURE VI on the third following page is a section of the form used to record activity in Receiving Inspection. In its entirety it includes all aspects of the operation.

Tool Inspection, another distinct service group within the division, had its own peculiar set of conditions which affected its budget performance. The records of hours spent (FIGURE V) and the equipment inspected enabled us to calculate the manhour requirements per general class of tool. Total manpower was then obtained from these figures and the forecasted workload. The Tool Inspection "scatter diagram" indicates costs for inspection as related to a total of the following unweighted Direct Labor figures:

FIGURE IV

LEAR, INCORPORATED | Figures Given Are For Illustration Only | **GRAND RAPIDS, MICH.**

VARIABLE OVERHEAD BUDGET PERFORMANCE

Department: Flex Shop Inspection **Department Head:** Stevhan

	WEEK OF: January 23, 1955			MONTH TO DATE		
	Budget	**Actual**	**Favorable (Unfavorable)**	**Budget**	**Actual**	**Favorable (Unfavorable)**
DIRECT PAYROLL						
Productive Run Time	$ 9731	$10157	$ (426)	$32,676	$31,898	$ 778
Set up And Reset (at ____ %)	564	524	40	1459	1526	(67)
Rework (at ____ %)	398	504	(106)	1232	1021	211
Service Repair (at ____ %)						
Pre Interim (at ____ %)						
Overhaul (at ____ %)						
Short Orders (at ____ %)						
Miscellaneous (at ____ %)						
Expensed D.L. (at ____ %)						
(Dept. Direct Labor Payroll) Total Activity Index	10,693	11,185	(492)	35,367	34,445	922
INDIRECT PAYROLL						
Variable (at 12.7 %)	1420	1216	204	4382	3913	469
Fixed	167	167	-	501	501	-
Sub-Total	1587	1383	204	4883	4414	469
Lost Time (at ____ %)						
Union Time (at ____ %)						
General Mach. Set Up (at ____ %)						
Holiday Pay						
Vacation Pay	165	140	25	330	140	190
Overtime Premium	138	151	(13)	441	552	(111)
Bonus						
Total Payroll	1890	1674	216	5654	5106	
EXPENSES						
Variable (at .5 %)						
Fixed Controllable						
Sub-Total						
Payroll Fringe Items	25	25	-	75	75	-
Tooling Costs						
Fixed Other						
Total Expense	25	25	-	75	75	-
TOTAL OVERHEAD COST	1915	1699	216	5729	5181	548

Budget Approval

DEPARTMENT HEAD DIVISION HEAD CONTROLLER General Mgr.

Figures Given Are For Illustration Only

1.   All Production Departments (Machine Shop, Gyro Labs, Electronic Assembly, E-M Assembly, etc.)

2.   Flex Shop

FIGURE V

NAME ____________________   DATE ____________________

TIME SPENT

	St. Time	1½ Time	2 Time
LT TOOLS			
T GAGE REINSPECTION			
USG REINSPECTION			
NEW VENDOR TOOLS			
NEW LEAR TOOLS			
HANGAR #1			
HANGAR #2			
GENERAL			
CLASSIFIED WORK			
OTHER (Specify)			

LEAR TOOLS RETURNED
FROM VENDOR ______

3.   Tool Room

The section has the responsibility of:

1.   Inspection of electronic test equipment prior to use as acceptance media.

2.   Calibration of this same equipment to see that it continues to accept only good parts and rejects bad.

3.   Inspection of standard purpose tools such as plug gages, ring gages, snap gages, etc.

4.   Production jig and fixture inspection.....in cases where the jig or fixture is used for final acceptance of the product.

5.   Inspection of mechanical gages on return to the crib after use.

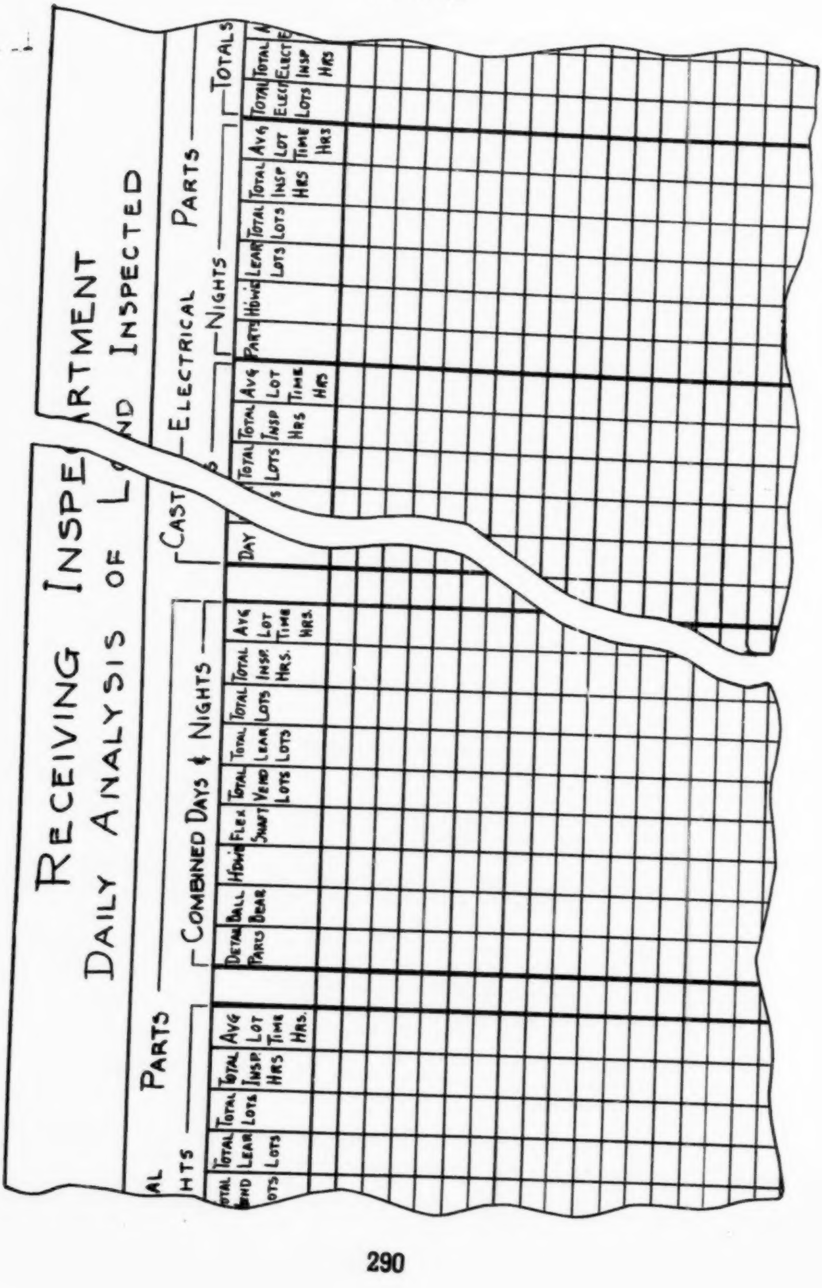
6.   Inspection of production gages and fixtures on return to the tool crib after a run has been completed.

7.   Calibration of electronic test equipment at the sub-contractors.

Also included is the acceptance of electronic test equipment destined for use by our customers in the maintenance of our products in the field.

Several further improvements are in work at the present time to

FIGURE VI



RECEIVING INSPE
ARTMENT
DAILY ANALYSIS OF L
ND INSPECTED
AL PARTS
HTS
COMBINED DAYS & NIGHTS
CAST
ELECTRICAL PARTS
NIGHTS
TOTALS
DETAIL PARTS
BALL BEAR
HDW'E
FLEX SHAFT
TOTAL VEND LOTS
TOTAL LEAR LOTS
TOTAL LOTS
TOTAL INSP. HRS.
AVG LOT TIME HRS.
DAY
PARTS
TOTAL ELECT LOTS
TOTAL ELECT INSP HRS

bring the present procedure more in line with the needs of the division. Some of these improvements are:

1. A more current issuance of expenditure reports.

2. A clearer allocation of costs into the proper accounts.

3. A more specific allocation of Fixed expenditures under Indirect Payroll.

4. The determination of an accurate activity index for premium pay costs.

5. The establishment of a schedule for periodic review of the accuracy of activity indexes.

The flexible budget with its mobility of use has proven in our case to be the most effective way to keep a close control of Q.C. costs. Its two distinct elements - (1) relatively current reporting, and (2) flexibility - enable a rapid recognition of danger areas and give clues to corrective measures. It is believed that many of the ideas set forth in this discussion are readily adaptable to other Q.C. operations, and their use should greatly assist in improving cost ratios.

QUALITY CONTROL TECHNIQUES USED IN THE FOOD INDUSTRY

Floyd J. Hosking
Corn Industries Research Foundation, Inc.

In order to determine the present extent and nature of statistical quality control used in the manufacture of food, a lengthy questionnaire was sent to more than 1,000 food manufacturers having over $1 million capital, as listed in "Thomas's Wholesale Grocer and Kindred Trade Register," 56th annual edition, published July, 1944.

The questionnaire used follows, and in each space is given the tabulated results obtained from 166 usable returns received up to February 28.

Some of the answers require comments. These are given below:

Comment A, question 5: Of the reporting 166 manufacturers, only 105 said they used statistical quality control. The approximate value of the food products made or shipped by the 76 firms using statistical quality control (29 companies did not report) was $4.6 billion. Of the 61 manufacturers who did not use statistical quality control, 55 (6 companies did not report) said the value of their food production in 1954 was $1.9 billion. Adjusting for nonreporters, the total value is nearly $8.5 billion. This is approximately one-fourth of the total value of food products shipped from plants in 1954.

Comment B, question 9: Prior to 1938, very little statistical quality control was used by food manufacturers, but from this date on, and especially after 1947, considerable use was made of statistical procedures.

Comment C, question 12: Other officials mentioned included (in order of number): General manager, production manager, technical director, factory manager, director or vice president of quality control, and director or vice president of research.

Comment D, question 13: The range of answers to this question was from 1 percent to 100 percent. The modal value was 100 percent, with the following percentages in order of frequency: 50 percent, 80 percent, and 25 percent. No answers were given by 16 manufacturers.

Comment E, question 14: Eleven answers, mostly in the "much" and "some" level of usage, were in the following miscellaneous uses of statistical quality control procedures: Troubleshooting, research, inventory reconciliation, sanitation, standards for machinery operators, machine development, and safety. "Research" was mentioned more than any other category.

Comment F, question 16: The following reasons were written in under "i" and "j" (in order of frequency): Improved analytical and sampling methods, improved equipment performance, location of equipment failures before they became serious, established new standards of quality, and increased efficiency of workers. No answers were given by five firms to question 16, a-j.

Comment G, question 17: While "production of higher quality products" was given as the most helpful feature of statistical quality

control procedures in 1954 by the food manufacturers, there were several comments under "i" and "j" including: Improved analytical and sampling methods, and improved equipment performance. No answers were given to this question by seven manufacturers.

Summary: A survey of 166 food manufacturers whose total output in 1954 was valued at nearly $8.5 billion reveals extensive use of statistical quality control. However, there is still considerable room for further use.

A Questionnaire on
Statistical Quality Control

Month Day Year

1. Your Company

2. Location of your main office
Street address City State

3. Number of your food plants 1,419 Location
State only

4. What foods were prepared or manufactured by your company in 1954? (1) (2)
(3) (4)
List in order of importance

5. What was the approximate value of all food products made or shipped by your company in 1954?
$ 6,535,000,000. (See comment A)

6. How many of your plants now have a quality control department? 672
Number

7. How many persons (full-time equivalent) were employed in your quality control departments in all of your plants in December, 1954? 3,449
Number

8. To what extent did your quality control departments employ statistical methods in 1954? Much (65%-100%) 25 Some (35%-65%) 34 Little (1%-35%) 46 None 61

9. In what year were statistical methods first used in any of your quality control departments? ____
Year
(See comment B)

10. Have you used statistical methods at any time in the past to a greater extent than indicated in Question 8?
8 Yes 143 No

11. If your company now uses statistical methods to a limited degree, or not at all, in connection with the quality control of your products is it because of: high cost 20 personnel shortage 30 not applicable 49 too formal 13 too difficult 12 too new 27; if there are other reasons, please specify lack of knowledge, 3; not useful enough, 2.

If your answer to Question 8 above was "NONE," you need not answer the questions below

12. To what top executive in your company is your statistical quality control department (or its

personnel) directly responsible (check one): President 20, Exec. V.P. 13, V.P. of Production 35, V.P. of Sales 1, V.P. of Purchasing 1; if another executive, please specify (See comment C).

13. What proportion of the food products (value basis) made or shipped from all of your plants in 1954 was subjected to statistical quality control techniques? (See comment D) Percent. (Give estimate if accurate data are not available)

14. In what department or areas of your company operations (all plants) did you use statistical quality control procedures in 1954? (Check in form below)

	Much	Some	Little	None
Purchasing - - - - - - - - - i.e., Conformance of materials purchased to your specifications	25	18	21	8
Manufacturing - - - - - - - i.e., Controlling your product quality	45	36	14	1
Packaging - - - - - - - - - i.e., Checking your weighing and filling machines	48	23	18	3
Other (give examples):				
(Also see comment E)				

15. Did statistical quality control save you money in 1954?

69 Yes   8 No   24 Don't know

16. In what way did statistical quality control methods help you in 1954?

79 (a) Produced higher quality products
56 (b) Reduced failures or rejections
24 (c) Reduced production delays
10 (d) Provided more prompt deliveries
67 (e) Gave better control of package weights
54 (f) Aided in meeting criteria of customers
32 (g) Reduced raw-material losses
28 (h) Improved business relations between your purchasing department and vendors
Other reasons (please specify):
____ (i) (Also see comment F)
____ (j) ____

17. Which of the points mentioned in question 16 represented the most outstanding contribution of statistical quality control to your company in any recent year?
(a) 57 (b) 15 (c) 8 (d) --- (e) 22
(f) 13 (g) 8 (h) 8 (i) ____ (j) ____
(See comment G)

18. What were the short-comings or faults of statistical quality control as conducted in your company? ________

____

____

19. Do you wish to receive a summary of answers to the above questions?

________ Yes ________ No

________________________ Your signature ________________ Title

____ Address

THE PURPOSE AND MEANING OF CONTROL CHARTS

C. C. Craig
University of Michigan

It may seem that my choice of subject is a bit peculiar because it is already pretty generally understood and there is no particular need for using up a period in this convention on it. It is true that there are quite a number of people who understand what I am going to talk about quite as well as I hope I do but my observation has led me to believe that there is a good proportion of SQC people who could benefit from a discussion of the basic principles and purposes of control charts.

I do not intend to catalog all the specific kinds of information about specific processes that good SQC men have read off of control charts illustrated by success stories. Of course in view of their essentially simple nature it is really quite surprising what an investigative tool control charts in imaginative and competent hands can be made to be. But that aspect of the subject of control charts has been elaborated upon in many books and still more articles in IQC and other journals.

Rather I want to discuss basic principles and purposes and try to emphasize in as clear a way as I can why I believe it is important to have a good grasp of them.

First, I believe it will be worthwhile to dwell at some length on what from conversations with Dr. Shewhart and from his writings I believe the man who invented control charts conceived their fundamental purpose to be. Second, and more briefly, I want to review what I among others have said and written before about the essential nature of control charts as statistical tools.

One can state the importance of having a process in statistical control quite succinctly: It is possible to make reliable predictions about the output of a controlled process but not about the output of an uncontrolled process. Once the meaning of the statistical control is understood this statement does cover the essential point provided also that its implications are understood. But I, at least, have been seriously disturbed for some time by the amount of evidence there is that the operational significance of statistical control in relation to predictability has somehow failed to get across to a rather large proportion of SQC practitioners. Apparently pretty commonly it has been felt or at least hoped that this matter is something for the professors to worry over but it need not bother the practical man too much. However, at least in quality control, some of the professors are concerned with matters of practical importance and this one on which the whole structure of SQC rests needs to be understood in order to assess its operational significance.

What is the importance for the practical man of getting a process into statistical control? This is simply as I said before, to make it possible to make reliable predictions about the output, either in the future or, from a sample, about what has already been produced. Of course we are all familiar with the fact that there is always variability in the quality characteristics of the product of any process and predictions or statements concerning the degree to which it will conform

to specifications have to be in the form of probability statements. But I am only echoing Shewhart when I assert that the sixty-four dollar question, which one neglects to consider at his own peril is: "Under what conditions can such predictions safely be made?"

You have all seen demonstrations using normal bowls or batches of colored beads in which by means of a succession of randomly drawn samples it is shown that it is possible to make valid inferences concerning the composition of the bowl or the batch of beads. I hope it was made clear that these inferences were made using the rules of probability and that they are not always, on every occasion, correct. They are valid only in the sense that we can control the percentages of cases in which they are correct. We can make this percentage high by not making our statements unduly exact in relation to the amount of evidence on which they are based.

That is, such demonstrations, unless they are rigged in a quite unrealistic fashion, can fail and they do sometimes, due simply to the workings of the laws of chance. This is of a piece with the fact that a sound sampling inspection plan, used in the most correct way, will always result in a certain percentage of incorrect decisions. But there are other very obvious ways in which sampling demonstrations could be made to fail. In my experience QC classes have always been made up of very cooperative and earnest people who give me and my demonstrations every break. But suppose we did get a few Peck's bad boys in a class. One of them behind somebody's back could remove half of the red beads from the tray; another could add a handful of red beads. Or somebody could easily pick out only chips with large numbers from the normal bowl; worse yet he could substitute an entirely different bowl for the one the instructor started around the class. Then the careful calculations according to the rules of probability made by the instructor would have no relation whatever to the results obtained.

You may remember that during World War II a lot of selling of SQC was done on the sweeping assertion that once it was installed one could immediately reduce his inspection force by a half or three quarters or even more because all inspection could be put on a sampling basis. There are plenty of places where the harm done by that statement still lingers.

Let us suppose that the XYZ Corporation makes widgets and it has ambitions to capture a larger share of the widget market. It authorizes a high powered advertising campaign and it hires a top designer to make its widgets more beautiful than any other make. And for the long pull and maybe because it has some ideals, management instructs the quality director to see to it that the quality of its widgets is kept up or even improved. It seems advisable to get a check on the current quality of the product and the next morning 100 widgets are inspected as they come off the line and four are found defective. The quality director realizes that 4% defective is probably not the true average quality and he calls in the statistician they have just hired from a university with an extra large stadium. This expert consults some tables and does some figuring and reports that the statement that the true average per cent defective lies between 1.6% and 9.9% has a 95% chance of being correct. The young man faithfully followed what he had been taught by a professor who never went to football games but his conclusion is a snare and a delusion. For the facts are that the entire sample was produced between eight and ten in the morning which is by all odds the best time to make widgets

though nobody has ever suspected it since the day's product has always been mixed together. It just happened, too, that the outside vendor who supplied components that went into this batch was considerably better than the other suppliers also regularly used for these parts. It also happened that the components built in the plant were produced from a better lot of stock than the average and the quality of stock was a lot more critical factor than anybody knew. Finally the assembly crew was the best one working in the plant, a fact believed by the crew members but proven to nobody else. I could also add that the inspector used felt that he ought to be careful not to make the reported lot quality any worse than it actually was. None of the various relevant circumstances were known because past inspection methods had never been designed in any way to let them show. At any rate inferences made from this one sample which was tied to some rather exceptional conditions, using the best textbook mathematics were worse than useless. And though a sample randomly selected from all of the company's product could give a pretty good estimate of the average quality being made it would not even suggest that there were realizable conditions under which better quality could be turned out at little increase in cost.

What is important to realize in this example is that if the production of widgets were to a constant average fraction defective all day long, from one shift to the next, from one time to the next, irrespective of the source of the component parts, with uniform inspection procedures, it would not matter at what time or place a sample was taken. So long as it was chosen randomly one could make an estimate of the process quality whose precision would depend only on the sample size. That is, if the process were in statistical control, one could apply the laws of chance to any random sample drawn from it.

Suppose that for a sampling experiment the class has not one but half a dozen trays of red and white beads to draw from. Then the most valuable demonstration one could put on, it seems to me, would be to show a way of taking samples and analyzing the results that would reveal differences among the composition of the trays either from tray to tray or from time to time. Shewhart's p chart with rational subgrouping was designed to do this job either for trays of beads or for the manufacture of widgets. The proper emphasis in statistical quality is not on the inspection of the finished product but on the inspection of the process. When a good doctor conducts a physical examination, he checks all of the important bodily functions that affect the patient's health and he takes steps to right those out of order if possible. Once the patient is in good health he likes to take periodic checks to see that the patient stays that way.

But my point has even more importance and I think it can be more clearly illustrated if we turn from inspection by attributes to inspection by variables. After all sampling theory for attributes inspection, either from large lots or from a process, is intrinsically simpler than for variables inspection. The underlying probability law for a process in control in the first case is always the binomial distribution which is specified by just one quantity, the process average fraction defective. But in the second case where the quality characteristic is measured, the underlying probability law for a controlled process depends on at least two quantities, the process center, usually denoted by $\bar{X}'$, and the process variability, ordinarily measured by σ'. Moreover we further have to specify the form of the distribution which we customarily take to be normal or at least approximately normal.

Now in the case of variables a process in statistical control does not have to obey a normal distribution law. It is not difficult to find examples of non-normality; there are plenty of instances in which one can be quite sure in advance that even if a stable cause system is in operation it will give rise to a skewed distribution or one that departs in other ways from the normal law. Actually nature has been lenient toward practicing statisticians in three important ways: We are very fortunate that such a high proportion of experimentally obtained distributions are at least approximately normal, that averages ordinarily tend so strongly to be normally distributed even if single observations do not, and that the commonly used statistical tests whose theory rests on assumed normality seem to be surprisingly insensitive to moderate departures from normality. It does seem, however, that awareness of these facts might lead one to wonder if it would not be possible to overdo one's trust in the kindness of nature.

Thus the importance of detecting that a controlled process does not obey a normal distribution law depends upon the circumstances. One familiar instance is again the case in which the product is the mixed output of more than one production line, of two or more shifts, of two or more parallel machines, etc. Here if the separate production units are each running in control with a normal distribution but around different centers, so long as each unit contributes a constant share to the total, the combined output will be in statistical control but not according to the normal law. As an example, suppose three normal bowls of 200 chips each with $\sigma' = 1.726$ for each bowl but with $\bar{X}' = -2, 0,$ and 2 respectively are all mixed together in one bowl. This new bowl is not normal but it is so nearly so that no random sampling from it will ever reveal its departure from normality. You may say, "Then why bother with this effort at detection"? My answer is that you should not bother unless the combined bowl's σ' of 2.376 is uncomfortably large for the tolerances that have to be met whereas a $\sigma' = 1.726$ would make things much easier. For emphasis I might add that even if the three bowl means were -3, 0, and 3 you still do not have a good chance of finding out that the bowl is a mixture by sampling it. Of course, sampling from the separate processes, that is by use of the principle of rational subgrouping with an $\bar{X}$ and R or an $\bar{X}$ and σ chart will quickly reveal process differences if they are anywhere near as large as for our bowls.

Let us grant, if only for the sake of argument, that it may be of real importance to get a process in which the quality characteristic is a variable into statistical control. Now such a process may be out of control for assignable causes that affect the stability of its center or of its variability or both. These assignable causes may be associated with different times, places, operators, machines, raw materials, vendors, inspectors, or even conditions hitherto unsuspected. How can the presence of assignable causes be shown and then localized so that one may hope to identify and cure them? The effective instrument Shewhart devised for this purpose is the control chart employing rational subgrouping. A frequency histogram of even a large sample taken with no particular design is generally so poor a means of checking for control that it often hides more information than it reveals. If a process is <u>known</u> to be in control then the only requirement of a sample is that it be random and a histogram can be used to estimate the process center and the process variability for now they exist. If the sample is large enough one can also learn if the underlying distribution is approximately normal or not. But the plain fact is that until control is estab-

lished the use of a frequency histogram analysis is very often logically equivalent to begging the really important question.

In spite of all I have said about the importance of getting a process into statistical control if one wants to make predictions concerning it, the situation that really led me to this talk arises in connection with sampling acceptance plans. As you know the essential information concerning the performance of any such plan is its operating characteristic (OC) curve. To install an acceptance sampling plan without knowing what proportion of lots at given quality levels it will accept is a really prime example of buying a pig in a poke.

Now in order to calculate an OC curve one assumes a succession of values for incoming lot qualities and in each case does a probability computation to find the chance that a lot of that quality will be accepted. If the inspection is by attributes, the quality is simply the number of defectives in the lot divided by the lot size. This is true no matter how little control there was in the process which produced the lot. The only restriction to insist on is that the sample be drawn randomly from the lot. The necessary calculations are straight forward and are only a matter of arithmetic. But when inspection is by variables the situation becomes more difficult. Examine the derivation of the existing standard variables acceptance plans. You will find that in every case they are built on the assumption that the distribution of the measurable quality characteristic in the lot is normal! If the lot is not too small and if it came from a controlled process obeying a normal law one can have confidence that the distribution in the lot will be approximately normal. The fact that the same per cent of defectives can arise from a whole set of combinations of process center and process variability adds a very considerable complexity to the computation of a point on the OC curve but we at least have a definite problem we can get hold of and results can be obtained. But if the process from which the lot came is not in control there is no necessity for the distribution in the lot to be even a ninth cousin to a normal one. There is simply no way to proceed using criteria based on variables to calculate the chance that a lot will be passed unless one has reasonably definite information concerning the distribution pattern in the lot, and, to repeat, for lots coming from an uncontrolled process that kind of information one does not have. It is true that OC curves do exist for variables sampling plans but they are only for processes or more strictly lots that obey stable normal distribution laws but with different $\bar{X}$'s and σ's which produce different proportions of the output which are acceptable on the specifications. It is also true that variables acceptance sampling plans customarily call for much smaller samples than attributes plans but there is a double penalty one must pay to get this greater sensitivity with safety. One is that one must take measurements, which may not be much of a penalty, and the other is that there must be at least a reasonable facsimile of process control in the production of the lots.

You may ask about the number of people who are getting good results with variables sampling plans. I can only say that if they have not checked on the state of control of the processes involved they are leaning heavily on providence to keep those processes on good behavior. It is true that if the test criterion is an average, as it usually is, nature gives them some protection there, too. There is a better question: If one has to have process control in order to make sound use of vari-

ables acceptance plans then what useful purpose do such plans serve? It is not a completely satisfactory answer to say that such plans can discriminate between the product of processes in control that meet specifications from the product of processes in control which do not meet specifications. But this question does not embarrass me at all for I maintain that the real purpose of SQC is process control, not the screening of unsatisfactory product.

This brings me back to what, I hope it is evident, is my main theme. The real purpose of the control chart is to get manufacturing processes in control. It is somewhat remarkable that the same instrument which furnishes the evidence that process is or is not in statistical control can also be so useful in locating the trouble spots which must be dealt with to establish control. It is a further valuable feature that once a chart indicates a state of control it has built in it estimates of the process characteristics on which reliable predictions of process performance can be made.

Now, finally, I want to briefly point out the essential feature of $\bar{X}$ and R or $\bar{X}$ and σ charts as statistical tools that makes them so effective as a means for studying the state of control of a manufacturing process. You have been well indoctrinated, I presume, with the idea that individual samples should be taken under as nearly constant conditions as possible. It is hoped that then the variability <u>within</u> samples, measured by either the range or the standard deviation, is at least a first approximation to the intrinsic process variability under a constant cause system. Now if the process is in control it will not matter when or where samples are taken so long as pieces to go in the sample are selected before it is known how the measurement to be made on them will turn out. A very important consequence of this is that the variability from one sample to another will then be due only to the same stable cause system that is operating within samples. Thus if we measure variability among samples by the variation among sample means, for a process in control, this variation will be strictly compatible with the within sample variability. The logical order is to first check whether the ranges (or sigmas) within samples show no more than chance fluctuation, i.e., see if the R's or σ's remain within their limit lines. If they do for at least 20 or 25 samples we assume that the process is in control with respect to variability. Then limit lines are set on the $\bar{X}$ chart in accordance with the estimated within sample variability. Finally if the $\bar{X}$'s remain within their limit lines the process is behaving like a controlled process. Of course there are precautions to be observed as to when and where samples shall be taken. This comes under the heading of rational subgrouping. The substantially correct rule to follow is to take the samples in such a way as to make the variation within samples as small as possible and the variation among samples as large as possible. If no matter how samples are spaced in time or in location or with respect to personnel or any other conditions of manufacture, the variation within samples shows only chance fluctuations, and the variation among samples remains in accordance with the within sample variation the process is in statistical control; it is obeying a probability law. But if there is a way of spacing so that the among sample variation is greater than that called for by the within sample variation, then something has been taking place between samples not explainable on the basis of chance alone.

I hope it is clearer to some of you by now that the essential feature of $\bar{X}$ and R (or σ) charts that gives them their great utility is

the comparison of within and among sample variation. I hope it is also clear that any device which really tests a going process for the maintenance of statistical control must contain in it an equivalent device.

In the case of fraction defective charts or of defects per unit charts the allowable among sample variation is directly set by the average quality level over the samples. They are thus simpler than variables charts but they are also much less sensitive and less useful in locating trouble.

The real heart of statistical quality control is process control. And for process control the control chart is a remarkably well designed and effective instrument. My advice to quality control people who are not using control charts is that they at least ought to know to what extent they are trusting providence to do their job for them.

A CUSTOMER'S PHILOSOPHY FOR QUALITY ASSURANCE

Colonel Clair A. Peterson
Headquarters, Air Materiel Command
Wright-Patterson Air Force Base, Ohio

TODAY, when all business activities are being so closely scrutinized from the viewpoint of management, it is surely high time that we - the members of the Government-Industry team - take an equally realistic view of the producer-customer relationship in "Quality Control."

In beginning such a study, perhaps the best way would be to repeat, word for word, the instructions given by the Air Service for the inspection of its first aircraft. The instructions go like this: "Start inspection at bulkhead just aft of cockpit and proceed right fuselage to empennage, around empennage, up left side of fuselage, around left wing, around engine."

Was this the actual beginning of "surveillance"? Was this "100% inspection"? Was this "eyeball measurement" just a method of determining whether or not the customer was happy with the product?

Well, these questions could be answered in various ways. And possibly, to a certain degree, any or all of the answers might be considered correct. However, one thing is certain: the customer (the Air Service) was definitely interested in finding out just what "quality" of aircraft was being actually delivered.

Inspection functions grew; the number of inspectors, both in Government and in Industry, rapidly increased. Before long, with world events serving as the creator of necessity, there were thousands of Government Inspectors (actually 14,000 at one time in the AAF) performing "supervision" over contractors' production systems. When 100% inspection became burdensome, the practice of "supervision over" came into being. It was not long, though, until "supervision over" gave way (early in 1943) to the much more satisfactory concept of "surveillance." Now let us examine, briefly but carefully, the essential principles of this "surveillance" concept.

We find first of all that the basic AF Quality Control Policy states explicitly that each Government Quality Control representative is responsible for carrying out his functions to assure conformance to these fundamental requirements:

First: Conformance to contractual requirements of supplies presented to the Air Force shall be determined on the basis of objective quality evidence. Such evidence will be obtained by the contractor, and will be evaluated and verified by the Quality Control representative exercising surveillance over the contractor's facility. Evidence may also be obtained independently by AF Quality Control personnel.

Second: Product inspection by AF Quality Control personnel will be used to the extent necessary to verify evidence of quality submitted by the contractor, or it may be used to determine acceptability of supplies on an individual or lot basis.

Third: The amount of evidence obtained or verified through product inspection by AF Quality Control personnel will depend upon the nature and the intended use of the product, and the effectiveness of the contractor's control over quality.

This is a customer's basic policy for obtaining quality satisfaction. It need not be looked upon as something truly "GI": it is, in fact, a written expression of what all of us, as customers, are actually doing today. We in the Air Force have ventured into the realm of the specific: declaring openly what we want, what we can do, and what we are willing to do, in our relations with our many producers.

Three definite phases of activity are of the utmost importance in properly exercising this "surveillance" Quality Control policy:

First: Detection

Second: Prevention

Third: Data Feed-back

Actually, all of us can carry out, as customers, the first objective, "detection." If, in spite of every precaution, a defect is found, we may either accept or reject the item. It is merely a matter of weighing the conformance evidence.

To practice "prevention," on the other hand, would be difficult for us, as customers, because normally it would be beyond our scope. And as we think of prevention as being practically synonymous with "control," obviously it becomes the producer's responsibility to control his processes properly, in order to prevent defectiveness. To protect himself, the customer may demand evidence, or assurance of control, from the producer. However, by doing this the customer is actually contributing to the control within the producer's plant.

"Data Feed-back" is not a controversial subject. Every producer is of course anxious to know how well his product is doing in the field - how it is going over with the customer. In other words, how is it selling? Data feed-back systems must be clear, concise, free of red-tape, responsive and timely. The results provided should be received in time for the producer to do something about everything requiring action.

We of the Air Force Quality Control organization must take an active part in all of the three objectives which make surveillance possible, because:

First: We must provide equitable treatment to all producers.

Second: We must provide protection for the taxpayers' dollars.

Third: We must provide coordination of contractual matters between the Government and the producer.

Equitable treatment of producers includes not only protecting the rights of all citizens to compete fairly for Government contracts, but it also includes assuring that producers deliver both quality and quantity in accordance with the contract. Not the least of the many advantages inherent in Government Quality Control is the fact that it definitely protects the quality producer from marginal and "fly-by-night" operators. Furthermore, if the contractual quality standards were not enforced, producers of inferior products would eventually drive responsible producers into a corner - for the pressures of competition are indeed many!

Since World War II, there has been a progressive development and an ever-expanding recognition of the "Surveillance Type of Customer Quality Control." Unfortunately, until recently much misunderstanding has existed, regarding both the nature of Surveillance Quality Control programs and their impact upon Industry. Perhaps this misunderstanding might have been more accurately described as criticism, for we have to admit that at first the concept of surveillance was ill-defined.

Now let us look at a typical Quality Assurance function - one of the many that must be performed by the customer. In the first place the customer must know what he wants: not just in quantity, but in the combined attributes which constitute quality. Furthermore, if the customer is buying an item which was designed and manufactured by a particular industry, then the customer must recognize that the producer also has the responsibility for controlling the quality of the item produced. The customer may conduct tests (wear, use, or operate the item), he may take the articles to a private laboratory for test, or he may decide to use his own "know-how" to accept the item. This can be interpreted as customer surveillance of the end item.

However, let us take another case. A typical milk processor advertises, "Don't buy unless you get the best - visit our plant - see our Quality Control." Well, the customer does just that. In fact, he may see several milk-processing plants. Assuming that the chemical and mineral content is up to required standards, the customer will probably buy products from the producer who shows positive, factory-floor evidence that the workers are doing a sanitary, wholesome, and quality-wise job. We might call this a customer surveillance of the processes. However, if (a few quarts or a few months later) the customer should find a hair in one bottle, he might revisit the customer's plant, or he might change over to another company, or he might consider this a mere chance and go on with the company as long as he is satisfied.

Under either of the two examples cited, there are of course many "if's, and's and but's." So let's explain further. Company "A" is a contractor; but it is also a customer, since it has hundreds of vendors or subcontractors. The Air Force is strictly a customer. Now, logically, neither Company "A" nor the Air Force can be everywhere, on a 100% basis, to see that every item conforms to its individual specifications. So let's face it. We are both customers, looking for the right item - that is, customer satisfaction. For practical economic and psychological reasons, we must resort to surveillance (less than 100% product inspection on each and every item) methods, ways and means. In order to attain satisfaction, we must use scientific techniques for

evaluating and verifying quality evidence. The test records, systems, reports, and inspection certificates which our producers create by controlling the quality and by inspecting and testing their own products all become a part of the surveillance program.

In some instances we, as customers, find it necessary to verify some of the evidence by visiting the producer, and then re-inspecting or re-testing the item in question, or perhaps by witnessing these functions and comparing the results. However, in most cases it can be confidently assumed that the more positively the item conforms to contractual requirements, the less surveillance we have to give the systems, procedures, and techniques utilized by the producer.

We the Air Force, as one of the world's largest customers, cannot possibly match inspector for inspector with the producer. Neither can you as a customer match inspector for inspector with your producers. Actually, it is inevitable that we as customers are buying, not just a physical article: we are also buying a service - the service of having the producer's Quality Control system assure that the article conforms to the purchase order or to contractual requirements.

I have attempted to present a realistic look into the growth of this new industrial science - Surveillance Quality Control; and I have also tried to justify the need for using "surveillance" techniques, instead of "Policeman, catch-me-if-you-can" programs of matching inspector for inspector. This concept of the producer's responsibility for monitoring control over the quality of his product is so important that I'd like to go into the subject a little further.

We can agree that regardless of what product is being produced, its quality depends on the degree of control exercised in the various steps and processes of its manufacture. In other words, I am saying that the measure of the quality of the product is its conformance to the specification. Granting that all good managers faithfully exercise the basic principles of management, we can go into specific aspects which can and do affect the quality of a product. The most important of these aspects include:

(1) The budget-dollars allocated for procurement and manufacturing.

(2) The purchasing agent's ability to "buy right": not merely to get the right price, but also to get the right material.

(3) The producer-vendor relationship.

(4) The engineering and change processes.

(5) The material handling procedures.

(6) The scheduling and machine loading practices.

(7) General housekeeping and plant engineering.

(8) The condition and the accuracy of tools and gages used for manufacturing, and also those used for determining conformance of the article to the specification or blueprint.

We could go on and on - but I'm certain that every good producer recognizes that these factors (and many others) do support one of our basic principles: "Quality cannot be inspected into a product - it must be built in."

Yes, quality must be built in - but just what is this "quality" that we are building into our products? Of course it must be "good" quality. With no intention of being unduly dogmatic, let us examine "good." Webster states that "good" means (1) adapted to the end in view; (2) suited to its purpose; (3) of satisfactory quality. Here "quality" comes into view again. Webster states that "quality is that which distinguishes one person or thing from others; as color, weight, skill, characteristics - such as degree of excellence." Now we can get into the problem of defining "goodness" or "acceptability," from a consumer's point of view.

General Simon of the U. S. Army Ordnance, pointed out some years ago that two specifications are necessary to define a "good" product: the design specification and the acceptance specification. In everyday practice, we as customers do not make this distinction; but this point does serve to clarify our thinking, as well as to call attention to some headaches that bedevil both Industry and the Government.

The design specification should establish a goal, by defining what a product should be like: the acceptance specification should tell us how to measure the degree to which that goal is achieved. A "good" product, therefore, is described in the acceptance specification: it tells us how to arrive at a decision as to whether or not a product is acceptable to us, even though the product may not be perfect. The individual quality characteristics must be identified. Items to be inspected or tested must be separated from those which either need not or cannot be inspected or tested. We as customers know that every quality characteristic cannot be tested or inspected. This would be impractical, either physically or economically; so we as customers are forced to select those characteristics which are likely to give us the most accurate information about a particular product. The acceptance specification must also establish risks - the calculated statistical risks - that we may reasonably take in product evaluation. In other words, the acceptance specification must include definite sampling requirements; and it must also indicate how these measurements are to be made physically.

How, then, do we define a "good" product? The answer is that acceptable quality can be defined in terms of a given number of observations of well-defined quality characteristics, using a specific type of instrumentation. This may sound a little academic; but actually these are facts of practical value to anyone who spends time, effort, and money in inspection or testing.

If all the information we have mentioned is established in specifications or in standardized systems and procedures, then both the producer and the customer have the assurance that any decision on quality is objective. Objective evidence, therefore, eliminates the need for continuous duplication of effort within the producers' facilities, and also by the various customers. This is a measurable economy factor. In addition, objective quality evidence close to the machines prevents

excessive quality variability; at the acceptance end of the production line and at the beginning end of the customer's using line. The old proverb you see of "an ounce of prevention": Well, it is much wiser to take cognizance in process, than to sit in on a post-mortem after the product has been completed. And certainly you and I, as customers, do not like post-mortems. We become very unhappy trying to get our dollars' worth out of something that just won't run - or just won't do the job.

We as customers can evaluate the quality of a particular purchased item in various ways. However, we should strive always to stick to logical thinking and reasoning. We cannot take it for granted that one event is the direct result of the event that immediately preceded it. Basically, if we make a decision of this sort, then we have succumbed to one of the most common fallacies of logic - the "post hoc" fallacy. It is true that event one may affect event two; but on the other hand it may be part of a process which includes several causes. They may act and react similarly, and it may be difficult to tell which is the cause and which is the effect. It may be much more difficult to isolate the other factors. We must therefore continue the search, and collect facts: real proof must be demonstrated. Surveillance quality control makes possible the collection of real facts. In other words, decisions can then be made without jumping to false conclusions based upon the end item only.

To sum up as briefly as possible, the surveillance type of Quality Control has been developed and adapted to the procurement of Air Force items. The basis of this Surveillance Quality Control is the agreement that the contractor will always assume and exercise complete and independent responsibility for controlling the quality of supplies, as well as responsibility for their production and delivery. This agreement requires that the producer comply with all the provisions of the specifications: inspection, testing, and quality control requirements. These specifications are considered solely as producer requirements; and they are the basic instruments for implementing the surveillance concept. By surveillance techniques we, the Air Force customer, audit the producer's processes, systems, and procedures, instead of sitting side by side with him, sorting out or segregating the unsatisfactory items from those which are acceptable.

The real justification of Surveillance Quality Control rests in the quality of the items produced. The net results are more serviceable products, produced in minimum time, at minimum cost, delivered on time - and a satisfied customer. What more could we, as customers with common objectives, ask of American producers?

As one of the largest customers in the world, we of the Air Force take particular pride in our recorded history of 46 years of dealing with American producers, and acknowledging our customer satisfaction. We have been an alert customer, having devoted approximately 35 years to intensive inspection practices in one form or another, with great concentration on purely technical ability. During the past 11 years, however, we have made a definite shift from the techniques of inspection to the concept of surveillance quality control. This concept has proved to be the most effective method of customer quality assurance. The philosophy underlying this concept has brought about standardized

terminology, improved methods and techniques. We in the Air Force are proud to be an effective partner in the furtherance of this technological progress and in the establishment of Quality Control as a recognized Industrial Science.

DISCOVERY SAMPLING

Ervin F. Taylor
North American Aviation, Inc.

The Discovery Sampling by attributes technique is a totally new approach to the inspection sampling problem. The basic theory of this method was developed in 1950 by James R. Crawford of the Lockheed Aircraft Corporation. Since that time many refinements and adaptations have been made. It is a tribute to the original theory that so much versatility is possible.

Discovery Sampling is an inspection tool developed at the request of shop inspection. Unlike most other attribute sampling methods, Discovery Sample takes into consideration the sampling factors found in practice.

A SIMPLE DISCOVERY SAMPLING APPLICATION

One of the best methods of introducing Discovery Sampling is to give an example of the simplest application.

Assume the following conditions: a production area manufacturing similar types of product in lots fairly consistant in size under 1000 pieces, and on which 100% inspection is now being performed. A typical Discovery Sampling installation instruction to the inspector might be:

1. Select a sample of 10 items at random from the lot.
2. Inspect all items in the sample.
3. If no defectives are discovered in the sample, accept the lot.
4. If any defectives are discovered in the sample, screen the lot.

This procedure can assure an AOQL of less than 0.005.

The conditions set forth above may not exist. Sampling may have been used before, the types of product may differ radically or, the lot sizes may be quite large or quite small. These, and many other conditions, are factors for which compensation can be made to deliver a sampling system tailor-made for a specific application.

WHY DISCOVERY SAMPLING WORKS

The small sample sizes and low AOQL's of Discovery Sampling are a departure from those experienced with most other sampling plans. This is true because of three practical factors taken into consideration by Discovery Sampling.

First, there are only three types of lots presented to inspection: 1) 100% good lots, 2) 100% bad lots and 3) partially defective lots. The 100% good lots will be accepted and the 100% bad lots, rejected by any sampling plan. Therefore, the entire sampling risk is contained only in the partially defective lots. Furthermore, empirically it is known that the partially defective lots constitute a minority percentage of the total lots inspected.

Second, a small fraction defective is more likely to occur than a large fraction defective. In other words, most of the lots presented to inspection are good. Empirically and logically this is true, since no manufacturer could long stay in business if large fractions defective were equally likely to occur.

The partially defective lots presented to inspection form a frequency distribution similar to that shown in Fig. 1. The shape of this distribution is reasonably stable, and the distribution parameter is practically constant in most areas over long periods of time.

Third and last, the usual process average calculation (the number of defectives found divided by the number of items inspected) is all too often a false indication of the relative quality level. This can cause an unnecessary sampling plan adjustment. For example, the last ten lots inspected usually used for the process average calculation often consist of mostly 100% good lots, with very few partially defective and 100% bad lots. The amount of defectiveness is thus spread over the entire group of lots in the process average calculation.

A sampling plan, regardless of the type, is doing the best it can when it rejects 100% bad lots and accepts 100% good lots. A sampling plan is only in error on partially defective lots. The more variation in the number of partially defective lots presented to inspection, the oftener will the sampling plan be wrong and should be adjusted.

Since only 100% good lots and partially defective lots reach the stockroom, a more efficient process average measure is the percentage of partially defective lots delivered to stock. This process average is a satisfactory measure of the quality level even when used on a variety of product from a group of men and machines.

In summary, the three major factors forming the basis of Discovery Sampling are:

1. The entire sampling risk is contained in the partially defective lots.

2. A small fraction defective occurs more frequently than does a large fraction defective.

3. The percentage of partially defective lots delivered to stock is a satisfactory process average measure.

The AOQL in the simplest case is given by the equation:

$$AOQL = \frac{A}{4n} \qquad (1)$$

A COMPLEX DISCOVERY SAMPLING APPLICATION

An excellent method of visualizing a Discovery Sampling application with many of it's variations is to study a complex example.

Through the receiving inspection area of North American Aviation, Columbus, pass thousands of different items from standard bolts to complex hydraulic equipment.

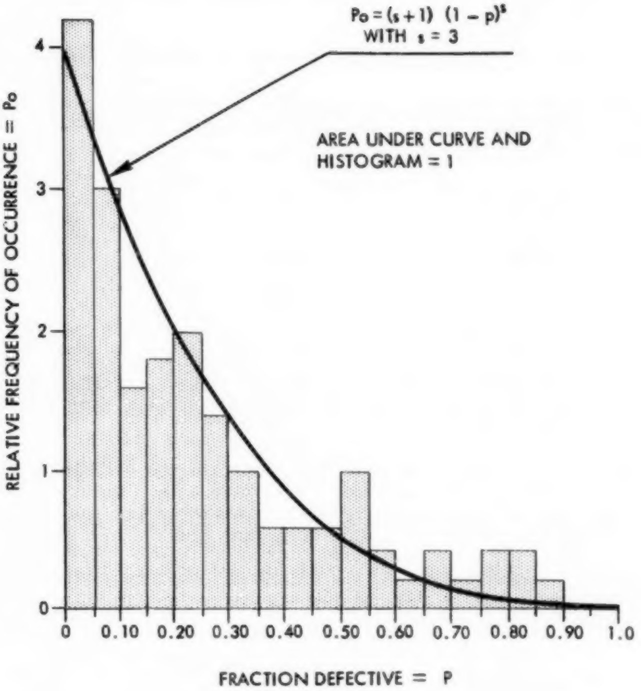


Figure 1 - Distribution of 100 Partially Defective Lots Compared with P_0 Curve

These items are received in lot sizes from 1 to more than 50,000. A classification of characteristics is used with a 4.0% AQL for minor, 1 5% for major and 0% for critical characteristics. All of the parts are not of the same general type and are received at widely varying intervals. Some sampling had been used with complete records being available for all sampling results. The first step was to make the following policy decisions:

1. Establish a general AOQL of 0.005.

2. Use individually controlled sampling plans for each area of application.

3. Vary the sample size weekly with the process average.

Several other problems required solutions before a workable system could be installed. The following problems were eliminated by the system:

1. Some samples are 100% good, yet the lots are partially defective.

2. Screening of lots is uneconomical.

3. Too much sampling for small lots.

4. Not enough sampling for large lots.

The second phase of the installation involved determining the extent of partially defective lots, the average past sample size and the new sample size.

The extent of partially defective lots and the average sample size were determined by examining past inspection data. It was easily determined that the probability distribution parameter was reasonably close to s 3 and was stable over wide areas and time.

A factor compensating for the amount of sampling performed was built into the equation for the new Discovery Sampling sample size. This equation is as follows:

$$n = \frac{A'}{4(AOQL)} \cdot \frac{(s+n'+1)}{(s+n'+1)-B(s+1)} \qquad (2)$$

where n = new sample size

n'= old sample size

A'= reported fraction of partially defective lots

B = fraction of lots sampled

s = partially defective lots distribution parameter = 3

AOQL = 0.005

The theory leading to Eq. (2) is given in the appendix to this paper.

If the true value for A is known, the following variation of Eq. (1) can be used in lieu of Eq. (2).

$$n = \frac{A}{4(AOQL)} \qquad (3)$$

Eq. (2) is unwieldy for shop use, therefore two nomographs, (Figs. 2 and 3) were designed.

Fig. 2 solves the equations:

$$A' = \frac{(\text{Partially Defective Lots})}{(\text{Partially Defective Lots}) + (100\%\ \text{Good Lots})} \qquad (4)$$

and

$$B = \frac{\text{Lots Sampled}}{\text{Lots Inspected}} \qquad (5)$$

Fig. 3 solves Eq. (2). No nomograph is necessary for Eq. (3).

The use of the charts is self-explanatory and has greatly simplified the calculation effort.

The sample size obtained from the nomograph is called the "Normal Sample Size". For small lots the reduced sample sizes in Tables I are used. For large lots (lots over 1000 pieces or more than twice the usual lot size) the "Normal Sample Size" is doubled.

The problem of screening defective lots was solved in this manner. If any defectives appear in the normal sample size, or equivalent for large or small lots, a defective lot has been discovered. The question of "how defective" is answered by taking an evaluation sample of enough additional pieces for a total sample of 100 pieces. Here, defectives are allowed, depending upon the minimum quality standard required. Table II gives AQL's for various acceptance numbers and a sample size of 100.

The Discovery Sampling Flow Chart Fig. 4 presents a graphic picture of the actual operation of the sampling plan. It is recommended that this type of chart be used to introduce Discovery Sampling to an area.

To summarize the 10 steps necessary to install a Discovery Sampling program in an area:

1. Check the general shape of the partially defective lot distribution from past inspection data.

2. Collect data on the number of partially defective lots, 100% good lots, total lots sampled, total lots inspected and average sample size from past inspection data.

3. Decide on an AOQL, and AQL's for each characteristic if a classification of characteristics is used.

4. Compute the normal sample size. If the AOQL = 0.005 the nomographs (Figs. 2 and 3) can be used.

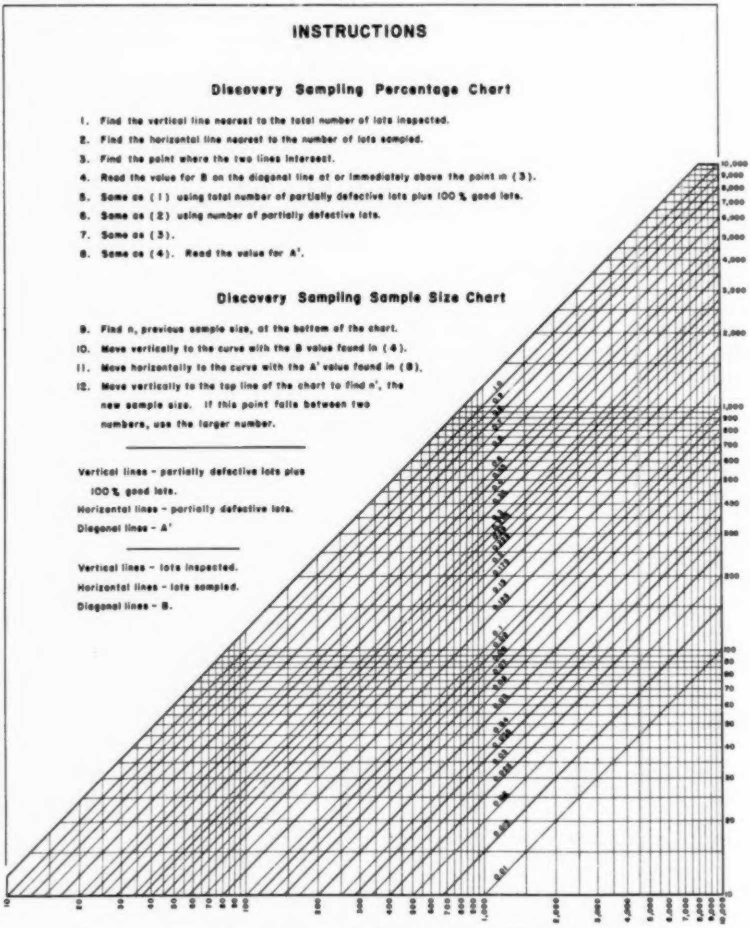


Figure 2 – Discovery Sampling Percentage Chart

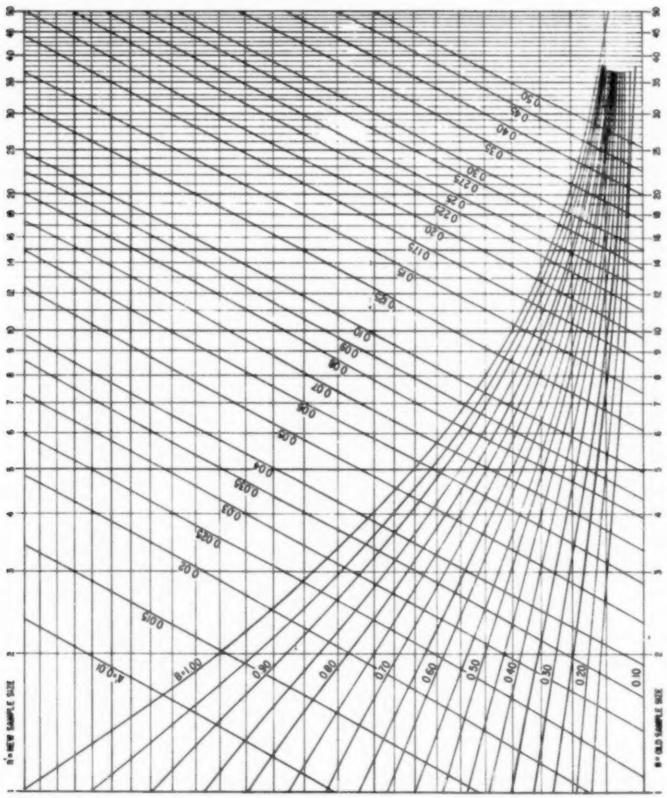


Figure 3—Discovery Sampling Sample Size Chart

TABLE I – Reduced Sample Size Table

REDUCED SAMPLE SIZE	NORMAL SAMPLE SIZE																			
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1-2	1-2	1-2	1-2	1-4	1-10	1-UP
2	2	2	2	2	2	2	2-3	2	2-3	2-3	2-3	2-3	2-3	3	3-4	3-5	3-7	5-19	11-UP	
3	3	3	3-4	3-4	3-4	3-4	4	3-4	4	4	4-5	4-5	4-5	4-6	5-8	6-11	8-30	20-UP		
4	4-5	4-5	5	5	5	5	5	5-6	5-6	5-6	6-8	6-8	6-9	7-11	9-16	12-43	31-UP			
5	6	6	6	6-7	6-7	6-7	6-7	7-8	7-8	7-9	8-10	9-12	10-15	12-22	17-59	44-UP				
6	7-8	7-8	7-8	8	8-9	8	8-10	9-10	9-11	10-13	11-15	13-19	16-28	23-76	60-UP					
7	9	9-10	9-10	9-10	10-11	9-12	11-12	11-14	12-15	14-18	16-23	20-33	29-95	77-UP						
8	10-11	11-12	11-12	11-13	12-14	13-15	13-16	15-18	16-22	19-29	24-43	36-116	96-UP							
9	12-14	13-14	13-15	14-16	15-17	16-19	17-22	19-26	23-33	29-51	44-139	117-UP								
10	15-16	15-17	16-18	17-20	18-23	20-25	23-36	27-39	34-60	52-165	140-UP									
11	17-20	18-21	19-23	21-25	23-29	26-35	37-45	40-71	61-192	166-UP										
12	21-24	22-26	24-29	26-33	30-39	36-52	46-80	72-221	193-UP											
13	25-29	27-32	30-37	34-43	40-58	53-91	81-252	222-UP												
14	30-36	33-41	38-50	46-66	59-102	92-286	253-UP													
15	37-46	42-56	51-73	67-115	103-321	287-UP														
16	47-61	57-81	74-97	116-358	322-UP															
17	62-90	82-141	98-397	359-UP																
18	91-155	142-439	398-UP																	
19	156-482	440-UP																		
20	483-UP																			

TABLE II

Relationship Between Defectives Allowed per 100 and AQL

Defectives Allowed per 100	AQL Producer's Risk 5%
0	0.1%
1	0.3
2	0.7
3	1.3
4	2.0
5	2.6
6	3.3
7	4.0
8	4.8
9	5.5
10	6.2
11	7.0
12	7.8
13	8.6
14	9.5
15	10.3

Computed from Reference (2).

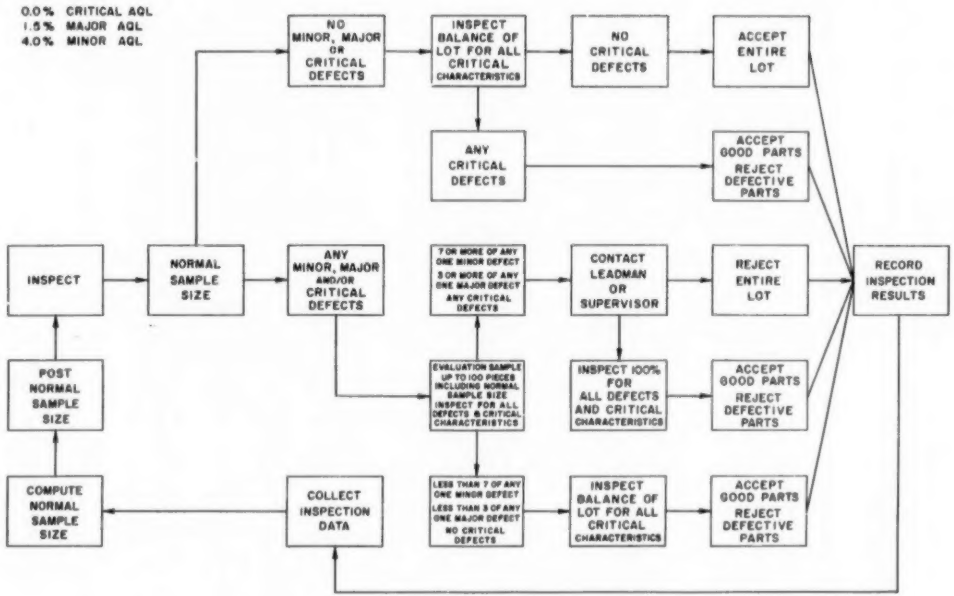


Figure 4 — Discovery Sampling Flow Chart

5. Establish some means of collecting sampling data. A tally of the number of partially defective lots, 100% good lots, number of lots sampled and number of lots inspected is all that is necessary.

6. Design a flow chart similar to Fig. 4.

7. Train inspectors on use of the plan.

8. Operate plan for a period of time, usually one week is an optimum interval.

9. Collect data at end of the period and compute a new sample size.

10. Post the sample size in the inspection area for the following period.

SUMMARY

The Discovery Sampling technique as described here is a powerful tool which can reduce the amount of inspection and inspection paperwork, increase the detection of defective lots of material and give a positive index of the quality of material passing into stock. It is a simple plan, easy to administer, operate and teach. The basic underlying theories are outlined in the Appendix to this paper.

The Discovery Sampling by attributes technique is the first in a series of new statistical tools. Other methods such as Discovery Sampling by variables, Decision Sampling with a fixed final run, band control chart, etc. are too lengthy to be included here. Future papers are planned for these techniques. Much progress is and will continue to be made in extending the basic theories of Discovery Sampling to cover other applications of sampling theory.

ACKNOWLEDGEMENTS

The Discovery Sampling theory was originated by James R. Crawford, Manager, Inspection Technical Services, Lockheed Aircraft Corporation, Burbank, California, William H. Wahrhaftig, Verne S. Myers, Robert A. Schafer and Ben K. Gold of Mr. Crawford's staff contributed heavily to the development. Seymour Sherman assisted with the mathematical logic.

Further refinements, the nomographs and the OC curve tables were developed by Walter J. Huebner, Jr. and A. E. Salisbury of the Quality Control Analysis Section, North American Aviation, Inc., Columbus, Ohio.

APPENDIX

BASIC THEORY

Discovery Sampling considers the probability of occurrence of a partially defective lot, the actual distribution of partially defective lots; as well as the probability that a sampling plan will accept the lot. Consideration of these probabilities determines an average outgoing quality limit and facilitates the construction of OC curves where the probability of occurrance of a partially defective lot is considered.

Lots presented for acceptance fall into three mutually exclusive classes.

Class	Symbol	Fraction Defective
100% Effective	(GDL)	$p = 0$
Partially Defective	(PDL)	$0 < p < 1$
100% Defective	(none)	$p = 1$

A lot which is 100% defective does not constitute a sampling risk; since it will be discovered if only one item is inspected. Hence, these lots will be excluded from further consideration.

The probability that a partially defective lot will occur is defined as: the ratio of the number of partially defective lots to the number of partially defective lots plus the number of 100% effective lots which are presented for acceptance during a given interval of time. Symbolically:

$$A = \frac{\Sigma_t(PDL)}{\Sigma_t(PDL) + \Sigma_t(GDL)} \tag{1}$$

A study was made to determine the distribution of partially defective lots. The probability density function

$$f(p;s) = (s+1)(1-p)^s dp; \quad s \geq 0, \quad 0<p<1 \tag{2}$$

was found to represent this data on a conservative basis. The values of the parameter "s" determined from the data were approximately 3. *

The probability of occurrence of a partially defective lot with fraction defective p may now be defined as:

$$P_o = A(s+1)(1-p)^s dp \tag{3}$$

The probability that a lot with fraction defective p will be accepted by a sample of size "n" with no defectives allowed is approximately:

$$P_a = (1-p)^n$$

* The question arises, would other studies also give this distribution? The writer has made many studies of partially defective lots. In every case Eq. (2) was applicable, although at times very conservatively.

Therefore, the probability that a lot with fraction defective p will both occur and be accepted is:

$$P_o P_a = A(s+1)(1-p)^{s+n}\, dp \qquad (4)$$

Assume that for each lot of size k there exists a set of lots of size k which have the distribution defined by (2). Then the following is true for each set of lots and hence for all sets of lots.

A lot with fraction defective p contributes to outgoing quality the fraction defective:

$$FD = A(s+1)\, p(1-p)^{s+n}\, dp$$

And the total fraction defective contributed to outgoing quality for all partially defective lots is:

$$\Sigma_p FD = A(s+1)\int_0^1 p(1-p)^{s+n}\, dp = A\frac{s+1}{(s+n+1)(s+n+2)} \qquad (5)$$

(Screening of lots in which defectives are found is assumed.)

Similarly the total fraction effective contributed to outgoing quality for all partially defective lots is:

$$\Sigma_p FE = A\left[1-(s+1)\int_0^1 (1-p)^s p\, dp\right] = A\frac{s+1}{s+2}$$

Also the total fraction effective contributed to outgoing quality for all 100% effective lots is:

$$\Sigma FE = 1 - A$$

The average outgoing quality may now be defined:

$$AOQ = \frac{\Sigma_p FD}{\Sigma FE + \Sigma_p FE + \Sigma_p FD}$$

or

$$AOQ = \frac{A(s+1)(s+2)}{(s+2-A)(s+n+1)(s+n+2) + A(s+1)(s+2)}$$

Considering AOQ as a function of s it is found that AOQ has a maximum value. Thus we may define the average outgoing quality limit (AOQL)

$$AOQL = \frac{A}{4n}\;, \quad A \leq \frac{1}{2} \qquad (6)$$

In equation (1) it was assumed that the true value of A was known. However, if sampling was applied, this is not the case. Let "B" denote the fraction of lots to which sampling was applied. Then from (4) it follows that:

$$BA(s+1)\int_0^1 (1-p)^{s+n'}\, dp = BA\frac{s+1}{s+n'+1}$$

(where n' = size of sample actually used.)

is the fraction of partially defective lots which would have been considered 100% effective lots. Thus if A' is the reported value of A then:

$$A' = A - AB\frac{s+1}{s+n'+1}$$

or

$$A = A'\frac{(s+n'+1)}{(s+n'+1)-B(s+1)}$$

Therefore (6) can be written

$$n = \frac{A'}{4(AOQL)} \cdot \frac{(s+n'+1)}{(s+n'+1)-B(s+1)} \qquad (7)$$

From equation (7) when n', B, A' and s are known n can be determined in order to insure any AOQL.

CUMULATIVE O. C. CURVES

From equation (4) it follows that the probability of acceptance for a lot with fraction defective p' or less is:

$$P_a(p \leq p') = A(s+1)\int_0^{p}(1-p)^{s+n}\,dp = \frac{A(s+1)}{s+n+1}\left\{1+(1-p)^{s+n+1}\right\}$$

Table III gives factors which facilitate the construction of these curves. Figure V gives an example of the use of this table.

SMALL LOT THEORY

Assume that for each lot which contains h items there exists a set of lots each containing h items which have the distribution defined by (2). From one of these lots with fraction defective p, a small lot (a lot which contains k items, with $k \leq h$) is taken. The probability that this lot will contain c defectives is:

$$P_c = \frac{k!}{(k-c)!\,c!}\,(1-p)^{k-c}\,p^c \qquad (8)$$

From (3) and (8) it follows that the probability that a lot with fraction defective p will occur and that a small lot which is taken from this lot will contain exactly c defectives is:

$$P_o\,P_c = \frac{A(s+1)\,k!}{(k-c)!\,c!}\int_0^1 (1-p)^{s+k+c}\,p^c\,dp$$

or

$$P_o\,P_c = \frac{A(s+1)k!\,(k-c+s)!}{(k-c)!\,(k+s+1)!}$$

If a lot containing k items with c defective items is sampled to the extent r with no defectives allowed in r, then the probability of acceptance (3) of such a lot is:

$$P_a = \frac{(k-c)!\,(k-r)!}{k!\,(k-c-r)!}$$

TABLE III

Factors for Computing Discovery Sampling OC Curves

N	p = 0	p = 0.05	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5	p = 1.0
1	0.000	0.050	0.100	0.200	0.300	0.400	0.500	1.000
2	0.000	0.049	0.095	0.180	0.255	0.320	0.375	0.500
3	0.000	0.048	0.090	0.149	0.219	0.261	0.292	0.333
4	0.000	0.046	0.088	0.147	0.190	0.218	0.234	0.250
5	0.000	0.045	0.082	0.134	0.166	0.184	0.194	0.200
6	0.000	0.044	0.078	0.123	0.147	0.159	0.164	0.167
7	0.000	0.043	0.075	0.113	0.131	0.139	0.142	0.143
8	0.000	0.042	0.071	0.104	0.118	0.123	0.125	0.125
9	0.000	0.041	0.068	0.096	0.107	0.110	0.111	0.111
10	0.000	0.040	0.065	0.089	0.097	0.099	0.100	0.100
11	0.000	0.039	0.062	0.083	0.089	0.091	0.091	0.091
12	0.000	0.038	0.060	0.078	0.082	0.083	0.083	0.083
13	0.000	0.037	0.057	0.073	0.076	0.077	0.077	0.077
14	0.000	0.037	0.055	0.068	0.071	0.071	0.071	0.071
15	0.000	0.036	0.053	0.064	0.066	0.067	0.067	0.067
16	0.000	0.035	0.051	0.061	0.062	0.063	0.063	0.063
17	0.000	0.034	0.049	0.058	0.059	0.059	0.059	0.059
18	0.000	0.034	0.047	0.055	0.055	0.056	0.056	0.056
19	0.000	0.033	0.046	0.052	0.053	0.053	0.053	0.053
20	0.000	0.032	0.044	0.049	0.050	0.050	0.050	0.050
21	0.000	0.031	0.042	0.047	0.048	0.048	0.048	0.048
22	0.000	0.031	0.041	0.045	0.045	0.045	0.045	0.045
23	0.000	0.030	0.040	0.043	0.043	0.043	0.043	0.043
24	0.000	0.030	0.038	0.041	0.042	0.042	0.042	0.042
25	0.000	0.029	0.037	0.040	0.040	0.040	0.040	0.040

With $N = s + n + 1$ and $P_a(p \leq p') = A(s+1)C_N$, where C_N is the value in the Nth row and the column corresponding to p'. (Figure 5 illustrates the use of this table.)

Assume that $A = 0.1$, $s = 3$, $n = 10$. Then $N = s+n+1 = 3+10+1 = 14$ and $s+1 = 3+1 = 4$.

Hence:

$Pa\ (p \leq 0.00) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.000) = 0.000$

$Pa\ (p \leq 0.05) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.037) = 0.0148$

$Pa\ (p \leq 0.10) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.055) = 0.0220$

$Pa\ (p \leq 0.20) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.068) = 0.0272$

$Pa\ (p \leq 0.30) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.071) = 0.0284$

$Pa\ (p \leq 0.40) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.071) = 0.0284$

$Pa\ (p \leq 0.50) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.071) = 0.0284$

$Pa\ (p \leq 1.00) = A\ (s+1)\ C_N = (0.1)\ (4)\ (0.071) = 0.0284$

Plotting these points and drawing a smooth curve through them gives:

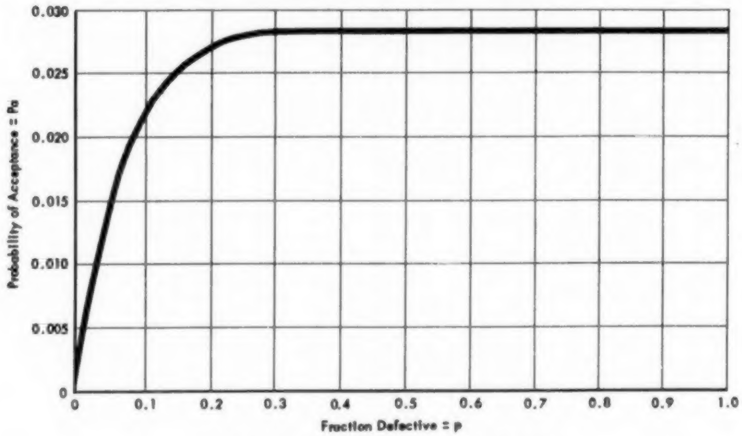


Figure 5 - Discovery Sampling OC Curve Calculation.

Such a lot has fraction defective c/k and the fraction defective contributed to outgoing quality by this lot is:

$$FD_c = \frac{c}{k} \cdot \frac{A(s+1)(k-r)!(k-c+s)!}{(k-c-r)!(k+s+1)!}$$

With s, k and r fixed, the average outgoing fraction defective for all c is:

$$\Sigma_c FD = A(s+1)\sum_{c=1}^{k-r} \frac{c}{k} \cdot \frac{(k-r)!\ (k-c+s)!}{(k-c-r)!(k+s+1)!}$$

By the use of factorial polynomials (4) this equation can be written in the form:

$$\Sigma_c FD = \frac{A(s+1)(k-r)}{k(r+s+1)(r+s+2)}$$

However, it follows from the basic theory that if the preceeding assumptions are fulfilled, then there exists an n which establishes any AOQL, and for which the fraction defective contributed to the outgoing quality for all partially defective lots (assuming that the lots are screened) is:

$$\Sigma_p FD = \frac{A(s+1)}{(s+n+1)(s+n+2)}$$

Thus, there exists a range of small lots for which:

$$\Sigma_c FD = \Sigma_p FD$$

or

$$\frac{k-r}{k(r+s+1)(r+s+2)} = \frac{1}{(n+s+1)(n+s+2)} \qquad (9)$$

From equation (9) the table (Table II) of reduced sample sizes was prepared.

LARGE LOT APPLICATIONS

The basic theory of Discovery Sampling is independent of the lot size. However, it is advantageous to decrease the probability of accepting an unusually large lot with a large fraction defective; and thus decrease the possible fluctuation in the AOQL due to the acceptance of an unusually large lot with a large fraction defective. In order to accomplish this and to maintain the simplicity of the sampling plan, it was decided that a lot with more than twice the usual number of pieces would be regarded as a large lot. The probability of acceptance of such a lot would be decreased by simply doubling the normal sample size.

DISCOVERY SAMPLING WITH A FIXED SAMPLE SIZE

If no assignable causes for fluctuations of A (Eq. 1) are present, then values of A may be assumed to be normally distributed. Since n bears a linear relationship to A (Eq. 6), values of n may also be assumed to be normally distributed.

An $\bar{X}$ and R (or X and R) chart can then be developed for n. (An $\bar{X}$ and R chart using a moving sample of four weekly values for n may be most convenient). If the $\bar{X}$ (or X) and R charts are in control after sufficient data have been collected, the weekly sample size can be replaced by a fixed sample size. This fixed sample size is:

$$n = \bar{n} + 3\sigma_n \qquad (10)$$

This provides Discovery Sampling with the practical advantage of an adequately conservative fixed sample size.

The fixed sample size determined by Eq. (10) is then posted in the inspection area. If a weekly plotting of the $\bar{X}$ (or X) or R chart indicates an out-of-control condition, the sample sizes are again posted weekly. Once control has been reestablished, a new fixed sample size is computed from Eq. (10) and posted.

REFERENCES

1. Crawford, James R., <u>Discovery Sampling</u>, Lockheed Aircraft Corp., Burbank, Calif., 1952.

2. Romig, Harry C., <u>50 - 100 Binomial tables</u>, John Wiley & Sons, Inc., New York, N. Y., 1953.

3. Fry. Probability and its Engineering Uses, D. Van Nostrand Company, Inc., New York, N. Y., 1928, p. 68.

4. Milne, W. E., Numerical Calculus, Princeton University Press, Princeton, New Jersey, 1949, p. 145 et seq.

USE OF STATISTICAL QUALITY CONTROL CHARTS ON CONTINUOUS PROCESSES

J. C. Dickson
Humble Oil and Refining Company

The problem of controlling the quality of products from large process units is one which must be constantly faced by the petroleum refiner and anyone else who operates continuous processes. One of a number of methods which can be used to help control these units is Statistical Quality Control. The practicality of using standard control chart techniques has been examined in an attempt to reduce quality variations at a minimum cost. It is the purpose of this paper to present some of the things which we have learned about using control charts on continuous processes and to explain in some detail one application which has been successfully carried out.

Most of the techniques which have been developed by the batch or parts industries can be applied successfully to the continuous process with only a moderate amount of modification. We cannot predict in advance just how successful a control chart will be, but we can identify some of the types of trouble which a control chart can help. Perhaps the most important source of quality variation which can be cured with a control chart is overcontrol, i.e., a process is adjusted more often or more severely than is necessary. Another use for the control chart has been the location of causes of variation with the consequent elimination of or compensation for these causes. The control chart also has value in the elimination of irrelevant specifications and inconsistent specifications on products.

Space does not permit covering all types of problems nor does it permit covering all the considerations which must be given to a problem before setting up a quality control program. However, much can be gained by following in some detail a specific application which has been made. For the purpose of this paper we shall call this problem "the quality control of a pipe still distillate stream".

The particular quality which was being controlled is not important here but perhaps the manner of controlling the quality is pertinent. The quality was changed by changing the withdrawal rate of the distillate stream under control or else by the withdrawal rates of other distillate streams at the pipe still. Obviously, the yields of the various products made from the pipe still are affected by how closely this quality can be controlled. This in turn produces the incentive for improving control, namely, a high yield of the more valuable distillates.

To initiate the quality control program we hoped that the historical operating data would be satisfactory to design the quality control chart. We reviewed two years of operating history and from these two years selected two consecutive months which represented the longest run which we could find without known upsets to the operation. These two months gave us approximately two hundred inspections for the quality in which we were interested. These inspections were examined to determine whether they had the normal probability distribution or not and it was found that they did not. The standard approach of averaging test results was taken so that statistics using the normal distribution could be used. The standard deviation for this inspection as obtained for the individual tests was interesting and surprising. If we had assumed a normal distribution for the individual test results and set control limits for our

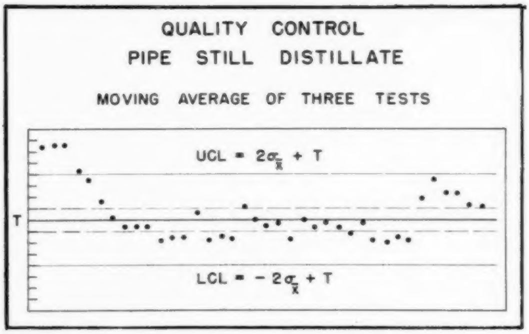
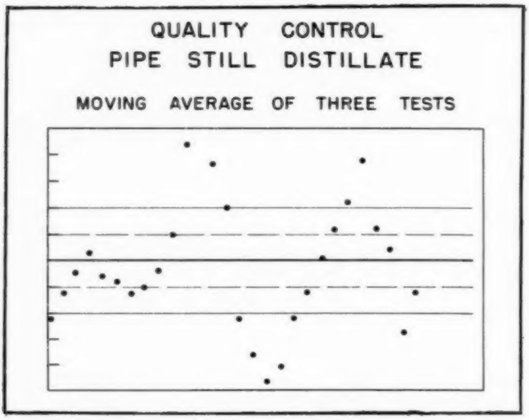


FIGURE 1

FIGURE 2



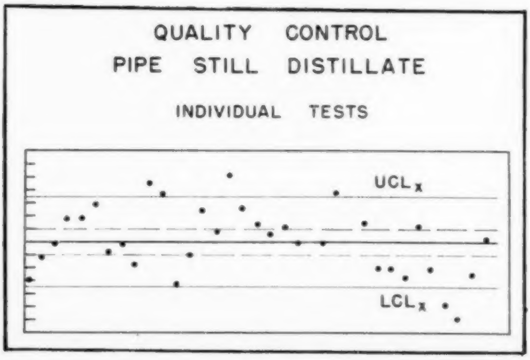


FIGURE 3

chart on the basis of individual results at plus and minus two standard deviations, our upper control limit would have been well above the theoretical maximum for the quality.

The testing of the distillate is both slow and expensive which makes replication impractical. Further, since hour-to-hour fluctuations within the process unit are as large as or larger than the testing error, replicate tests on the same sample would give little help. We, therefore, selected the moving average of three points as the factor to be plotted on the control chart. Figure 1 shows the first control chart which was installed at the pipe still. The dashed lines within the control limits represent the old "judgment" control limits which had been used for single tests. It is obvious that these limits were much too narrow for the process. It is also obvious that the new control limits were too wide for the variation experienced in the product quality. We, therefore, reviewed the more recent history of this quality, i.e., since the control chart was installed, and found a significant reduction in the standard deviation of this test inspection. We consequently revised the control limits and made them more narrow. This can be seen in Figure 2. The scale to the left has been changed but the magnitude of the change can be seen since the dashed lines in this figure again represent the old "judgment" basis control limits.

Shortly after this control chart was put into effect a severe cycling started as is obvious in the last two thirds of Figure 2. It will be well worth our time to examine this closely since it points out what I believe to be one of the most important points in the application of control charts to continuous processes.

The control chart shown in Figure 2 used the moving average of three tests as the control point. The moving average was selected to produce a normal distribution for the sample points and to give increased sensitivity to long-term or slow changes in the process level. The interval between individual tests was eight hours, which means that each point represents the average of the past twenty-four hours' operation. Furthermore, the testing of the sample required about four hours to complete. Therefore, the point plotted on the control chart was representative of an operation which is on the average sixteen hours in the past. This would be fine if you were mainly interested in where you had been but we are mostly interested in where we are going. Trying to control a process with sixteen-hour old readings is similar to trying to drive a car through the mountains by looking out the back window. It was this sort of thing which caused the cycling in Figure 2, although the severity of the cycle could have been substantially reduced through more extended operator training in the use of control charts.

The specific cause of the cycling shown in Figure 2 can be described as follows: An upset occurred which shifted the average quality level considerably above the target value for the quality, but since the first test result was averaged with two older tests obtained prior to the upset, the point did not go out of control immediately. Further, the range chart, which is not shown but which was in use, failed to show loss of control. However, after two tests were obtained after going out of control, the average was out of control and the operator adjusted the unit to correct for the trouble and another test was obtained. The most recent test was added to the two previous tests which were obtained during the poor operation and the average was again out of control so the operator made another adjustment. This, of course, was unnecessary. In

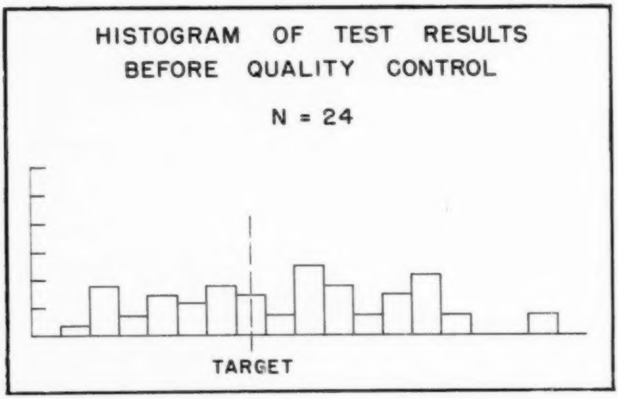


FIGURE 4

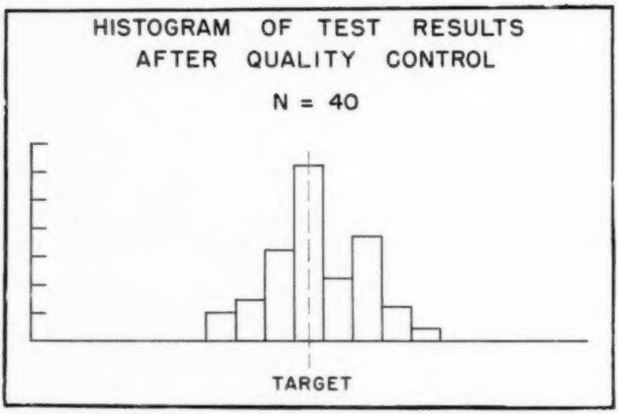


FIGURE 5

fact, the averaging of these three points was completely unjustified since the parent distributions are obviously different. But consider the alternatives. If the operator must obtain three new points before plotting a point on the control chart he must wait thirty-two hours before he has a point on his chart, and this is at a time when he needs the information more than at any other time. If he is to plot the individual point, he must have a special control chart for the purpose and it is not desirable to have a number of special charts with special rules at the process. It would only serve to confuse and not much else.

This led us to review the data again to see whether individual tests could now be used to control the process. It was found that they were satisfactory. The use of the control chart had reduced the quality variations to one which could easily be approximated by the normal distribution. We, therefore, set up the control chart shown in Figure 3 using individual test results rather than moving averages. Once again the dashed lines within the control limits represent the "judgment" basis control limits. There is still a high percentage of out of control points on this chart, but we have found that this is not only to be expected but also to be desired because a large amount of the variation which does occur in the quality of the products from a process can be controlled by good operators on the basis of laboratory test results. If there were no points out of control, it would indicate that the operators had little or nothing to do so far as adjusting the unit on the basis of the laboratory tests is concerned and we would reduce the testing frequency until the percentage of out of control points increased.

To illustrate the effect that the quality control chart has had upon the variance of product quality from this unit, you are referred to Figures 4 and 5. Figure 4 represents the histogram for the quality inspection which we obtained for the best week in the two months used for the original design. Figure 5 shows the histogram which we obtained during the first two weeks after lining out with the control chart for the individual tests. Nothing could be much more graphic than the comparison of these two figures.

The fact that statistical techniques are usually based on sampling error or a counting process does not rule out their application to continuous processes. We would perhaps like to think that the problems which we have with continuous processes are so difficult and so complex that the only thing to be done is to rely on history and just let things take their courses. This may be a salve to our conscience but it will never stuff our pocketbook. The continuous process has a lot in common with the processes of the parts industries. The persons responsible for quality control in the continuous process industries have a lot to learn from the parts industries today. And it is probably safe to say that a few years from now the quality control man in the parts industries will find that he has something to learn from the continuous processes.

HUMANIZED COMMUNICATIONS CAN OVERCOME RESISTANCE TO CHANGE

Ralph E. Burt
President of National "U" Association

In industry it takes change to make dollars. No plant can afford to stand still. Industry must achieve change, else change will be thrust upon the plant by competition and by the shifting influences upon plant people of environment, both inside and outside the plant.

Historically, our productive power has set the pace for material progress because we have been free to probe and to prove the potentials of change. Our economic expansion is rooted in the ability of the few to enthuse the many toward pioneering new horizons, new ideas, new inventions, new production methods, new markets, and new personal opportunities. But it was far easier to inspire people when frontiers were romantically geographic than it is to enthuse them today over the coldly scientific techniques which characterize the changes in modern production. The lure of free and fertile lands beyond the western hills had far more appeal for the individual than any of our little plans for simplifying jobs into a completely dull monotony.

In most instances today, plant leaders and workers must solve their problems right where they are. It is difficult to run away from human problems. We tend to take them with us, even when we move the plant or the family to what seem like greener pastures. It is not the weather outside the plant which determines the rate of resistance to change ... it is the human climate inside the plant. There is no escape from most of the problems attending changes in methods or machines short of reaching down within the feelings and attitudes of people just as they are in order to come up with the answers. Most of our plant problems are right within us ... within management people and production line people.

How can we overcome human resistance to change far more successfully than most of us are doing today? First of all, let us see exactly what is involved in these problems of change in relation to plant people. Let us examine the effects of change upon the attitudes of plant people and, therefore, upon their productivity and its Quality.

One does not have to visit Reno to learn that change is not confined to plant activity alone. Change is one of the fundamental essentials of life itself. Change has been going on constantly ever since Eve swapped her girlish confidence for a figleaf. The story of man is the story of change. One might guess, then, that man welcomes change with open arms. But plant leaders know this is not true.

Resistance to change is as constant a force as change itself. All living things are creatures of habit, clinging to old ways and hesitating to plunge into new and unaccustomed channels. Man is no exception. Men are always dreaming of greener pastures, but it takes some very diplomatic shoving to get them over the fence. If left to his own devices, man will risk new ways of thought and new modes of action only when he can summon up the enthusiasm to be confident of success. Without this enthusiasm, he will make but a half-hearted attempt to change and will resist changes thrust upon him. If we are to overcome this resistance, we must first learn how to supply and spark his enthusiasm for change.

The rapid succession of changes in our century has been largely in innovations very close to the heart of industrial growth. It is the modern changing production plant which is turning out the products and equipment behind and within our amazing advances in transportation and communication, in construction and destruction, and in a million and one appliances, gimmicks and gadgets. In this age of atoms and synthetics, manufacturing has replaced agriculture at the grass roots of our economic welfare and potential. Our frontiers lie not on the prairie but in the plant. But the question is: how awake are plant people to the significance of this truth in relation to their own work? Is this concept of industry too lofty for workers to care about or to comprehend?

There is no doubt but what plant people are aware of the wonderful things we now can do with our new machines, new laboratories and new methods. Outside the plant, supervisors and workers cannot help but be awake to the wonders of modern manufacture. What is more, the standards and importance of Quality are very clear to plant people when, as consumers off-the-job, they window-shop or buy what they need and can afford from a glittering array of exciting products. The well-dressed products from our production lines are filled with appeal and romance for all of us ... an appeal and romance that is heightened by masterful advertising and selling.

Now let us step inside the plant. Let us go into any one department where but one or two operations are taking place, and these on but a small part of the finished product. How much of the romance of that finished product ... how much of the excitement of our fabulous national production ... rubs off onto the worker or his immediate superior as they struggle to meet production schedules? Compared with the adventure found by the old-time craftsman in making a product from start to finish ... contrasted with the excitement which management and engineers derive from planning all of the operations essential to each finished product ... how much appeal and inspiration is there for most workers in turning out barrels of nuts for the left rear wheels of even the most deluxe automobiles? Just what is there to get enthusiastic about in the monotony which so often attends work simplification in practice? Are we justified in expecting a worker to sustain his appreciation for Quality standards as he punches out dozens of some part whose use and destination may be a mystery to him?

What is it that is adding glamor to the finished product? What is it that is making workers as off-the-job consumers alert to the importance of Quality? The answer should be evident to us all. Advertising and Salesmanship! Millions upon millions are being spent annually to convince these everyday consumers that they should save their pennies to enjoy the benefits of products made by other workers like themselves. It is a striking success story, this story of advertising and selling, and it is the dynamic and happy and positive force that is keeping so many plant wheels turning today.

Significantly, it is advertising and salesmanship which are helping us to maintain the accelerated pace of change at the consumer level. One almost might say we are being badgered into swapping for new models in nearly everything before we have fully grown used to the old style. But how much advertising and salesmanship is being used to accustom production workers to changes vital within their own work? How good a job of promotion are we doing within the plant to sell production workers and their immediate leaders more job enthusiasm? What portion of the plant

budget is devoted to making leaders and workers into plant customers for Quality? What efforts are being extended to glamorize and romanticize every job and every change in job methods? We have made excellent progress toward providing worthy leisure activities for workers, but what are we doing right on the job to arouse as much enthusiasm in a man for his work as our advertising and selling techniques have awakened in him for his off-the-job activities and comfort? With few exceptions, we are doing far less than an adequate job in advertising and selling enthusiasm for change to our own plant people.

This need for selling enthusiasm for change within the plant is all the more urgent today when machines and methods are changing almost before they can be bolted to the floor or blueprinted for action. It isn't the machine nor the method which does the grumbling and resisting in the face of these rapid changes. It's the plant people. And it is the plant people whom we must reach with the creativity and inspiration that will turn negative reaction into positive action. This is simple to say, but it is more difficult to do.

Very obviously, our first task is to build a better climate in which the ideas and the facts of changes will not run up against immovable walls of suspicion, indifference or dissension. Usually, such reactions are the outgrowth of confusion ... confusion largely in communications between management and workers, supervisors and operators. Some of this confusion is the inevitable outgrowth of any change from accustomed ways and habits. Father and Mother cannot install a home television set without upsetting old family habits, causing Dad to fall in the dark over the furniture Mother moved, and bringing quite a change in the behavior of Junior and his kid sister. Nor can new methods or machines be introduced within the plant without disturbing the feelings, thoughts and actions of the plant people involved. The best possible humanized communications, based on the best possible humanized understanding, obviously are demanded where plant people are inclined to believe that most changes are designed to speed up their production and company profits with little or no consideration for the sensitivity, safety and security of plant people as individual human beings.

On the other hand, plant people will receive the news of change in a far more optimistic, positive and cooperative way if they already are working in an atmosphere of mutual understanding and trust between management and workers. Nothing is more vital to a smooth and happy plant operation, and to the necessary introduction of important changes, than establishing right from the start a plant-wide climate of common understanding and mutual confidence in the basically good intentions of leaders and workers alike. We must share the conviction that all plant people, with only the fewest exceptions, at all plant levels are sincerely eager to do their jobs well. We must share in knowing that all plant people normally have a decent respect for each other's problems and purposes, but that circumstances can overpower that positive respect with doubts and fears and resentments, particularly between plant levels. When these negative elements are left to fester without the right answers and the right solutions, we are deliberately encouraging the growth of antagonistic attitudes which break out into stubborn hallucinations and bitter grievances. So it is increasingly imperative that we also share EMPATHY ... that remarkable individual ability to imagine oneself in the other fellow's shoes, facing his problems and hopes just as he has to face them.

None of us would question that our personal environments outside the

plant have done much to make us what we are. Our feeling and thinking upon matters of religion, ethics, politics, family problems, culture, social values and so on, all are the products of the way we've been brought up, the way we have lived, where and how and, maybe, with whom. We do not become entirely different people when we enter the plant, washing our brains of all of the virtues and faults that characterize each of us in his homelife. We do not lock our individual personalities outside the plant gate. And not even the most monotonous job will keep us from being ourselves, nor from wanting to be regarded as our individual selves in all communications and in all activities within the plant.

Knowing what tremendous influences are exercised upon each individual outside the plant by his own environment, how can any plant leader afford to ignore the opportunity to do a far better job in improving each individual's environment within the plant? Particularly his environment of ideas, for he must feel right and think right to work right! Knowing that outside the plant the great impact of all effective advertising and selling is aimed successfully at the individual, how can any plant leader put his entire confidence in communications within the plant that are mass media alone and fail to reach the individual at his own level, from his own point of view, and in his own production line language? Knowing that positive ideas cannot possibly flourish in a negative climate, how can plant leaders succumb to obvious petty politics or develop the superiority complex which seeks to interpret the individual without listening to him? Small wonder that we have misunderstanding, dissension and waste in public affairs when we are guilty of these same faults in private business!

The time to overcome resistance to change is before that resistance can break through like a poisonous weed to smother the plant. Even if the sourest reactions from changes already are upon us, let us move immediately to establish a positive climate of harmony for the future in order that the current dilemma will not diffuse itself into a chain reaction which will completely swamp us at the next crisis. But let us bear constantly in mind that a happy climate of confidence and cooperation cannot be maintained along the production line either by trying to <u>shame</u> employees into better work or by trying to <u>buy</u> their earnest efforts.

There is a tendency among Quality Control leaders to snatch at straws sometimes, rather than getting to the grass roots of this challenge of change. All of their eggs are put in the one basket of the change itself ... some new technical device, some new plan for charting, some new scheme for automation ... but the potential value of these changes, which may be excellent, often is ruined by their not preparing the way beforehand. Their positive plans are sabotaged by their own negligence in this matter of climate. Let us look at an example.

Today, there are plants whose Quality Control leaders are placing all of their hopes for better work from employees on charts which are hung at machines to record each operator's Quality progress, up or down. The first results often are positive, like the first swallow of raw vodka. But the novelty and profit from this system wears off. Leaders finally realize that it is always best to criticize in private but to praise in public. In planning these report cards, they might well have asked the schoolteachers who remember to publish the honor roll publicly but hand out the individual reports privately. They might have asked the schoolboy, lagging his way home to get Dad's signature on the bottom ... of his card, he hopes. The lad has learned that Dad will let him off with but a

stern speech if he but brings his marks to a passing grade. And the worker learns, unfortunately, to think in terms of getting by. He comes to do only what is necessary to meet his department's minimum standards. He is not inspired by this system to hit a real peak of performance, for he gets to believe that some workers just naturally excel him and personally sags into mediocrity.

The great teacher or great preacher may be able to inspire an individual to perform over his head but, if plant leaders were teachers or preachers ... if they were advertising experts or super-salesmen ... they would not be engineers and statisticians. In the area of human relations ... in meeting the challenge of selling job enthusiasm ... they need help. Help beyond what personnel departments are supplying today! That much needed help is available from those who sincerely concentrate on these particular problems dealing with the human side of plant people in relation to the objectives of your own leadership.

It is further true that money incentives alone will not stir and maintain lasting employee enthusiasm for Quality work. The index on wages and salaries notes a phenomenal rise throughout this era of modern production. But the chart on job enthusiasm generally indicates a curve that pitches downward to a dull thud. Measuring a man's worth by his wages alone is like deciding the value of a baby by the doctor's fee. Each of us has a keen interest in his personal income because each is aware of the rising demands from the family budget. Workers, like stockholders, enjoy seeing their income rise. But no one needs to elaborate on the number of negatives which have grown out of using the "Almighty Dollar" as the only way of giving a worker the feeling that he has a real status in the plant community. If all he asks is "what's in it for me", it is we who are letting the dollar factor outweigh the human factor in our own plant. It is impossible to buy a worker's enthusiasm. His interest, yes, but never his enthusiasm! And without that enthusiasm, he will always resist change, if only to make himself heard.

The chance to be heard ... indeed, the sacred right ... is at the grass roots of all of our education and all of our off-the-job environment. Inside the plant, are we stifling democracy? Or are we giving workers a full, free opportunity to release their job troubles before they become grievances?

What we urgently need today are effective TWO-WAY communications between management and workers. Without this two-way communication we shall never reach the peak of understanding, enthusiasm and performance. We must stop growing apart and start growing together if we are to conquer resistance to change. We cannot go forward with management believing that workers will never understand what a load plant leaders have to shoulder. Nor shall we go anywhere if workers are left without the means of blowing off friendly steam about their own job problems. We have to talk together, listen together, think together ... yes, and feel together ... and then we shall work together, no matter what changes become necessary.

Today, Quality Control leaders well can afford to take the bit in their teeth. Theirs is the chance to turn natural negatives in the present plant climate into dynamically natural positives. BY CHANGING THE CLIMATE. There is an easy, efficient and effective way to build plant-wide understanding of each individual's point of view. A plant-wide respect for common purposes. A plant-wide enthusiasm for change and for

progress. This will not happen without new and dynamic promotion. What direction will this promotion take?

In ancient languages one word often had two distinct meanings, according to its use. In Hebrew the word for "work" could also mean "reward", which in itself is something to think about. Today leaders in Inspection or Quality Control think of Quality as an objective. They forget its double-barreled meaning. Quality is more than a goal. Quality is first and foremost a language ... a language which can arouse every plant individual to think positively of common experiences, common attitudes, and common points of agreement which he shares with everyone in the plant from bottom to top. It is the one and only positive language for the free and harmonious discussion of change.

As consumers outside the plant, both leaders and workers are agreed on the importance of Quality in what they personally buy, each wanting the best for what he is able to pay. This desire for Quality is as positive within the union worker as it is within a pillar of the N.A.M. It is the same within a job supervisor as it is within a machine operator. It is easy and natural for all of us to understand the motives of plant customers in wanting work from us that meets their particular standards. And it is equally easy and natural to relate this customer attitude to the attitude we must have as producers and suppliers of these plant customers. There is no argument here, no dissension, no controversy, no prejudice, and no negatives. The Quality Language is dynamically positive for everyone!

Obviously, then, the plant program of communications on <u>every</u> job problem will be twice as effective if designed and delivered in the Quality Language. It is the one and only language which implies for each plant individual the personal opportunity to do something constructive about his own long range security without appealing for outside help. It is the one and only language which clarifies for the worker how he himself can boost steady sales, steady orders, and steady work through steady Quality. Only with the Quality Language can we arouse positive enthusiasm to meet the challenges of better work, safety, job housekeeping, steady attendance, and individual achievement with head, heart and hands.

Yours is a wonderful opportunity to humanize plant communications with the Quality Language. To build the climate in which changes will be welcomed at all plant levels. But applying the Quality Language demands from you the vivid, vital and vigorous use of the best advertising and selling techniques. Your plant can sell your products. But can you sell your people? Can you build the humanized communications to arouse job enthusiasm and Quality Enthusiasm? In the Quality Language and in the dialect of your production line!

From our own experiences in the National "U" Association ... from our own work with leading industries facing the problems of change and the challenge of Quality ... we urge you to develop a consistent, continuing and complete promotional program to meet the daily problems and hourly challenges of better human relations which alone can overcome resistance to change and indifference to Quality. Encourage the sincere efforts of other departments toward a better plant climate, but revitalize their work and your own with a program of showmanship and salesmanship in the Quality Language ... a program which your workers feel belongs to them, and is devoted to their personal needs, problems, responsibilities and opportunities. Human Relations is not a job for the Personnel Department

alone, nor should Quality be a challenge limited to your department alone. These are efforts of plant-wide significance, begging for the solid backing of management and of every plant leader.

In applying advertising and salesmanship to your program, recall the impact upon customers and prospects made by company trade-marks and themes which all of us recognize wherever we see them. If someone says to you, "the pause that refreshes", a particular beverage is uppermost in your mind. If a listening dog is pictured for you with the slogan, "his master's voice", another company comes quickly to mind. Establish a trade-mark and theme for your program to enlist Quality Enthusiasm as the best of all possible climates for plant progress. Make this a personal symbol and a personal theme. Stick with it and stick by it, as you stick by the trade-mark of your own firm.

Around this symbol and this theme, and bearing in mind the points we have made, build a program for your plant that always pictures and talks about those ideas and those things which already are familiar to your workers. Introduce your new ideas in settings that are popular and familiar with everyone whom you hope to convince. Above all, reach the individual. If you cannot inspire him, you will never budge the group of which he is a key part. Ask him to do what he likes to do, and thank him for doing what he thinks he does well. Doing these good things must be more than a part of plant policy ... they must be part of a well-organized program. Endow every step of your program with the element of the novel and unexpected, and your workers will accept change in the same friendly spirit as they welcome the features of your plant-wide promotion.

Notice, please, how we avoid the word, "campaign". It is a word which characterizes the conglomeration of fits and starts that seem to have a lot of uplift at first but customarily end in a sag. The chief value of these varied contests comes when they are projected within the closely related framework of your over-all program; that is, when you still preserve at all times the solid foundation of a human interest program with a continuing theme.

In action, the National "U" Association has found these steps effective when supported by the enthusiasm of plant leadership:

1. The harmonizing of directives from all department heads with the over-all plant program to create a positive climate with the Quality Language.

2. Regular meetings of representatives from all department heads with representatives from production workers, these latter representatives being the natural leaders among the workers, not their stewards or supervisors, since the purpose is not to discuss grievances but Quality problems as workers see them, helping workers to release these problems and to free themselves for better work.

3. Regular reports, including answers to Quality questions and solutions to Quality problems, channeled back to production workers by their rotating representatives.

4. The coordination of supervisory training with this democratic

system, designed to lighten the load of job leaders and to make their own appeals for good work more effective.

5. Enlistment of house organ editors in the Quality cause to bolster continuous backing up of the plant-wide promotion.

6. A thematic program carried all the way through into regular displays and distributions which give monthly emphasis to this plant program.

These steps may seem like a lot of work. The alternative can be a lot of trouble as you seek to bring your plant people into line with changing standards and methods. Actually, you will be amazed to realize how simple and easy a plan this program can prove to be for you, once you have it underway and are enjoying plant-wide cooperation. You will come to realize that your Quality promotion is as permanent a need as the plant's Safety program, and of the utmost importance to your customers.

As Quality leaders, yours is the challenge of the hour! Will your industry be able to cut production costs through eliminating poor work and by helping workers to feel individually responsible for doing every job right the first time? It's up to YOU. Will you be able to establish the harmonious climate of worker attitudes ... the happy TWO-WAY communication between management and workers ... that will have a positive impact upon every change instituted within your plant? It's up to YOU. Your work must be with men more than with machines, with hearts and hands more than with tools, with people more than with parts or products.

Over and over again, this has been proved: Quality is the language of success! It is the language that can meet every plant need. The language that can help YOU share a greater part in the progress of your plant. And so I urge you to wake up your own potentials with plant people ... to accept the challenge of Quality in these days of constant change ... and to enlist NOW an enthusiastic army of workers who want to make Quality a new symbol for personal and plant progress everywhere.

It's up to YOU!

QUALITY CONTROL AS AN ADMINISTRATIVE AID

Charles A. Bicking
Office of the Chief of Ordnance
Washington 25, D. C.

Introduction

The higher one looks in administrative levels of business, the more likely one is to find that decisions are based on tabular or graphic presentations of data. It is apparent that at the administrative level, at least, one of the principal methods of contact with the rest of the organization is through statistics. This is so well recognized that in some large organizations, "Chart Meetings" are a part of the regular routine at administrative levels.

What this means from the viewpoint of quality control is that when top administrators realize what it is all about, not only does quality control begin to roll on the production line but also it begins to find uses throughout the business structure.

Extent of Administrative Applications

The use of quality control methods to help solve administrative problems, therefore, has come about naturally. Actually, although administrative applications have not been as numerous nor as spectacular as those to production control, there has been a steady, parallel growth, even from the earliest days of statistical quality control. Over seventy literature references were listed during research for a recent paper(1) on management uses of statistics.

Clerical operations have offered some of the best opportunities for application. This is logical, because the products of clerical operations, although paper reports, are quite similar to manufactured products in that they lend themselves to sampling and charting techniques. It is remarkable to note the almost immediate improvement in quality of clerical operations after the installation of control charts to determine levels of performance. There can be little doubt that the application of quality control principles has brought about these improvements. This is true because when well designed, the control charts strip out differences between individuals or operations. Acceptable performance criteria are stated in advance of starting work. Reports are rendered, usually in the form of "p" charts, so that quality levels are known to all and intelligent decisions follow naturally. This is an improvement over the kind of situation which used to prevail in which the requirements of the job were not clearly explained to the workers and no systematic check was made to help the worker keep his work in line. Accounting records, inventory counts, or auditing records also lend themselves readily to quality control procedures. As a matter of fact, it is difficult to think of any activity of management that cannot be improved by some application of quality control techniques.

An Application to Inventory Control

One type of administrative application that has not received much publicity is the study of inventory control problems. Some time ago, the author was involved in a determination of optimum inventory size and of economic reorder points for a distribution operation of considerable

size. This operation involved twelve different products distributed from four warehousing points in the East and Middle West. A great deal of the study involved, of course, the enumerative approach ordinarily thought of as the domain of the business statistician. However, very critical parts of the analysis of the data were accomplished, in this instance, by the use of Shewhart control chart principles.

The products were bulk materials shipped in drums and accounting was made on the basis of the pounds shipped. Data were available for at least twelve months, in some instances for as many as 18 months, on the number of pounds of each of the twelve products shipped from each of the four warehouses.

The first column of figures in Table I shows the total actual monthly shipments of a typical product from one warehouse. Obviously, these figures vary so much that no sound statistical forecast could be made on the basis of these data alone. Actually, in the past, inventory levels had been established on the basis of accumulated forecasts by sales territories. This was generally unsatisfactory, however, partly because of the perennial optimism of salesman, partly because no really systematic use of data was being made to determine optimum inventories, and partly because warehouse districting was in need of revision. It had been common practice to tranship between warehouses or to ship across warehouse district lines to fill rush orders. As a result of these undesirable conditions, during the most recent period the over-all turnover had been only 3.3 times per year.

As a first step in the analysis, it was decided to reallocate all shipments for the period for which data had been accumulated on the basis of the logical shipments from each warehouse. A logical shipment was one which minimized the shipment cost and eliminated the need for transhipments from warehouse to warehouse.

In making this reallocation it was decided that a considerable portion of the total shipments, including all car lot shipments, could be made directly from the plant which was located near one of the larger centers of use, without intervening warehouse storage. The resulting logical shipments for the typical product used as an illustration are given in the second column of Table I.

The variation in logical shipments was much less than in the actual shipments and some statistical analysis seemed possible. Accordingly, for each set of data, control chart limits were computed based on the moving range of consecutive monthly figures. Typical range computations are given in the final column of Table I. Control chart limits for individuals and ranges were then computed, as follows:

Individual limits: $\bar{\bar{X}} \pm E_2 \ \bar{R}_M = 17{,}007 \pm 2.66 \times 6839 = 17{,}007 \pm 18{,}192 = 35{,}199$ and 0

Range limits: $D_4 \ \bar{R}_M$ and $D_3 \ \bar{R}_M = 3.267 \times 6839$ and $0 = 22{,}343$ and 0

These are very wide limits due to the variability still present in the estimates of logical shipments but they do indicate that it is not too unrealistic to use the moving range data to determine optimum inventories, at least on a preliminary basis. It was decided that in each case an inventory should be maintained capable of meeting 90% of the

Table I

Total Shipments in Pounds
Warehouse 1 - Product H

Month	Actual Shipments	Logical Shipments	Moving Range
July	16,650	11,700	
August	11,250	11,700	0
September	44,100	19,800	8,100
October	13,050	15,766	4,034
November	30,600	25,200	9,434
December	24,300	11,700	13,500
January	40,500	27,000	15,300
February	12,150	28,350	1,350
March	36,450	18,000	10,350
April	29,250	8,675	9,325
May	19,350	16,200	7,525
June	29,700	13,950	2,250
July	31,950	13,050	900
Totals	339,300	221,091	82,068
Averages	26,100	17,007	6,839

demands on the warehouse. The standard deviation of the logical shipment data from Table I is given by $s = \bar{R}_M/d_2 = 6839/1.128 = 6063$ with 12 degrees of freedom. An upper bound which will guarantee meeting 90% of the demands is given by $\bar{X} + ts = 17{,}007 + 1.78 \times 6063 = 17{,}007 + 10{,}792 = 27{,}799$.

This was rounded downward to 27,000 pounds which is considered to be the optimum inventory figure. This gives a turn-over rate of 7.5 times per year compared to the old over-all rate of 3.3. If there had been a marked trend in sales volume or if sales had fluctuated seasonally, it would have been necessary to apply appropriate corrections. However, this was not done and when the optimum inventory levels had been obtained as described for each material at each warehouse, the total optimum inventory at the four warehouses turned out to be 417,000 pounds compared to the previous peak inventory of 681,500 pounds, a reduction of 39%. The new turnover rate was estimated as 4.0 times per year compared to the old rate of 3.3. Furthermore, the total annual volume handled by the warehouses was reduced by 38% due to the practice of shipping to the local area or elsewhere by car lots direct from the plant.

Another desirable figure was the proper reorder point and this was arbitrarily set at the median logical monthly shipment size. Control charts, like the one for the example shown in Figure 1, portrayed all of the pertinent information graphically and assisted in the analysis of each situation.

No system may be devised which will work perfectly under all conditions and it was expected that this system would be subject to adjustment upward or downward for known seasonal fluctuations or for unusual changes in normal orders during a given month.

An Application to Rating of Technical Personnel

The 309 technical people employed in a research laboratory were assigned to one of nine groups, each under a group leader and from one to three assistant leaders. In applying a standard merit rating plan, the supervisory personnel were rated by the director of the laboratory and each group of researchers was rated by its own group leader or leaders. One final score, on a scale between 0 and 1200, was assigned to each individual. With this many raters involved in rating so many groups, many variations in the ratings are to be expected. One rater may rate all individuals in his group too high, another too low. A rater may be prejudiced and rate certain individuals high and others low. Furthermore, in any given group there may be individuals with qualities either far exceeding or much lower than those normal for the group or required for the type of work being done.

Comparisons may be made within each group or a standard for the whole laboratory may be determined and all groups compared to it. This latter is undoubtedly the best thing to do in this instance since the work of all the groups is very similar. However, as a first pass at analysis of results, the ratings for each group were arranged in subgroups of four and separate $\bar{X}$ and R limits were calculated for each group. Most of the groups were in pretty good control on this basis although the average rating and spread of limits differed widely from group to group. The control charts for one particularly well controlled group, Group N, are shown in Figure 2.

When it comes to choosing a standard basis for comparing ratings from all groups a little problem arises. What is a proper average level and what width of limits indicates a satisfactory rating job? It happened that the average for Group N, 719, was closest to the grand average score for all groups, 715. Furthermore, this group was an old, stable group with an experienced and very capable administrator as group leader. Also the limits, although narrowest of all the groups, represented a fair spread of scores and seemed to afford a good basis for distinguishing exceptionally good or exceptionally poor performance. The implication of failure of an average or range to stay within these limits would be that something unusual had affected the rating and that it should be investigated. If the cause could be traced back to an individual rating it might indicate the presence of an individual doing unsatisfactory work or having superior accomplishments and ready for promotion to a supervisory position. Action for training or better placement, or separation might be indicated. If evidence of some kind of bias on the part of the raters is indicated, the ratings of the suspected group would have to be examined in detail. Re-rating or rating by a different rater might be done.

When averages and ranges of all laboratory groups were plotted with limits based on the group N computations, a number of interesting results were observed. Although some of the groups were in statistical control, others showed lack of control on the average or on the range chart. Although the earlier examination of charts for single groups indicated that only a few individual ratings might be out of line, the condition of the combined charts indicated that differences between the raters was a major source of concern. As an example, the ratings by the leaders of Groups P, S, and T are compared in Figure 3 with the ratings by the leader of Group N. For Groups P and S the range chart is in control but the average chart shows a shift to the high side for Group P and shift to the low side for Group S. Presumably the leader of Group P has been too generous and the leader of Group S too harsh in their ratings. Both need further

training and practice in the use of the rating procedure. For Group T, however, the average rating is very close to the overall average and the sub-group averages are well controlled. The range chart is out of control. This indicates either a wide spread of abilities within his group or a tendency to rate some few individuals too severely and others too leniently. This rater also, should be called in for review of his rating technique.

One should possibly go slow in applying results of a control chart analysis of this sort without reservations. However, as a supplement to the routine methods for review of merit ratings this kind of analysis should prove very valuable in many instances.

An Application to Analysis of Indirect Expense

Comparison of one period or one operating unit with an other is the essence of control chart application to administrative problems. However, in a highly diversified operation it is sometimes difficult to find a measure which will be of the same order of magnitude for all segments of the operation. For this reason, percentages or ratios are often most useful for comparison purposes.

An example is found in a study of indirect expenses in which the measures used in a control chart analysis are selling expense, experimental expense, home office, branch office and indirect mill expense, and total indirect expenses, all expressed as per cent of sales. Data were available on an annual basis for six operating departments covering a period of eight years.

The experience of each department was divided into two sub-groups of four years each, as shown in Table II, for the data on selling expense in per cent of sales. Scanning of the table turns up the fact that there are many discrepancies which make direct comparison difficult. For example, throughout the first four year period, department S was way out of

Table II
Selling Expense in Percentage of Sales

	Department					
Year	C	E	N	P	S	V
1	2.55	11.40	8.91	6.11	26.54	0.63
2	2.47	9.99	7.37	6.57	33.03	0.39
3	4.07	10.83	10.07	10.06	22.78	0.79
4	3.25	9.54	9.34	8.47	12.55	0.77
Subgroup Average	3.09	10.44	8.92	7.80	23.73	0.65
Subgroup Range	1.60	1.86	2.70	3.95	20.48	0.40
5	3.02	8.05	10.67	8.46	12.60	0.54
6	2.65	4.17	10.07	6.32	9.31	0.36
7	2.29	3.05	9.77	6.76	7.43	0.25
8	3.13	10.99	7.77	6.24	7.80	0.72
Subgroup Average	2.77	6.56	9.57	6.95	9.29	0.47
Subgroup Range	0.84	7.94	2.90	2.22	5.17	0.47

line on the high side. This was explained by the fact this was a newly organized department which was not expected to pay its way during the early years. As a matter of fact, for the first four years, total

indirect expenses amounted to 50% of sales, although by the fourth year almost all classes of expense were almost down to normal. Since there was no doubt that this department was "out of control" during this period, that set of data was omitted from the control chart calculations. A question is raised also about Department V, which was always very low. There is an explanation for this, the fact that the department makes relatively few bulk products most of which are used by the other operating departments. Likewise, Department C is obviously different. Actually, it is an old department selling a small number of products to well established markets. Although such results may be desired as the ultimate goal of all departments, it did not seem reasonable to include them in determining limits in the current comparison.

Seven subgroups were used, therefore, to calculate limits for a chart of individual measurements, as follows:

$$\text{Individual limits: } \bar{\bar{X}} \pm E_2\bar{R} = 8.50 \pm 1.457 \times 3.82$$
$$= 8.50 \pm 5.57 = 14.07 \text{ and } 2.93$$
$$\text{Range limits: } D_4\bar{R} \text{ and } D_3\bar{R} = 2.282 \times 3.82 \text{ and } 0$$
$$= 8.72 \text{ and } 0$$

The resulting control charts are shown in Figure 4. Department C was out of control on the low side, as was to be expected. Department E while in control, was above the average during the first period and below average for most of the second. This department was most affected by military demands and the high results were certainly due to lack of military demand in those years. Department N was always above the average. This department was selling in a highly competitive market and had, except for the new department, S, the highest average selling expense for the whole eight years. This is not a desirable situation but may be an unavoidable one. It was noted that the experimental expenses of this department were also high. For department P, the overall control is very good. Department S was out of control on the high side during the early years. Department V was, as noted, always out of control on the low side.

A similar analysis was made for the three other measurements studied with much the same results.

The generally satisfactory experience of departments C and P are shown at a glance. The temporary difficulties of department S are very apparent while the unusual experience of department V stands out. The peculiar dependence of department E on military purchasing is clearly marked. The similarity between the competitive situations of departments N and S is particularly clear due to the similarity of results in the last four year period.

It is believed that this kind of an analysis is helpful to general management in comparing and understanding operations. Furthermore, this kind of information can be of use to the individual operating departments by leading them to action to keep their results in line.

Conclusion

Stated in general terms, the use of statistical and graphic techniques provides a method for reaching decisions and directing action to control costs. The further down the managerial ladder we go, the less organization we find for decision, action and cost control. We pay well

for administrative ability and for the statistical information on which administrative decisions are founded. Through a quality control program, the same advantages are obtained down to the lowest supervisory level on a self-paying basis. A highly respected administrative tool is extended in its scope and usefulness. Because it represents an extension of an essentially managerial function, it should be directed from a policy-making level. Since very often in industry administrators arise from the ranks, the extension of the appreciation of the value in statistics will provide a means of training for administrative responsibilities.

Reference

(1) Bicking, C. A. and Lorber, S. J., "Management Uses of Statistics" Steel Processing, 40, 2: 108-10, February 1954

LOGICAL MONTHLY SHIPMENTS - WAREHOUSE No. 1
PRODUCT H

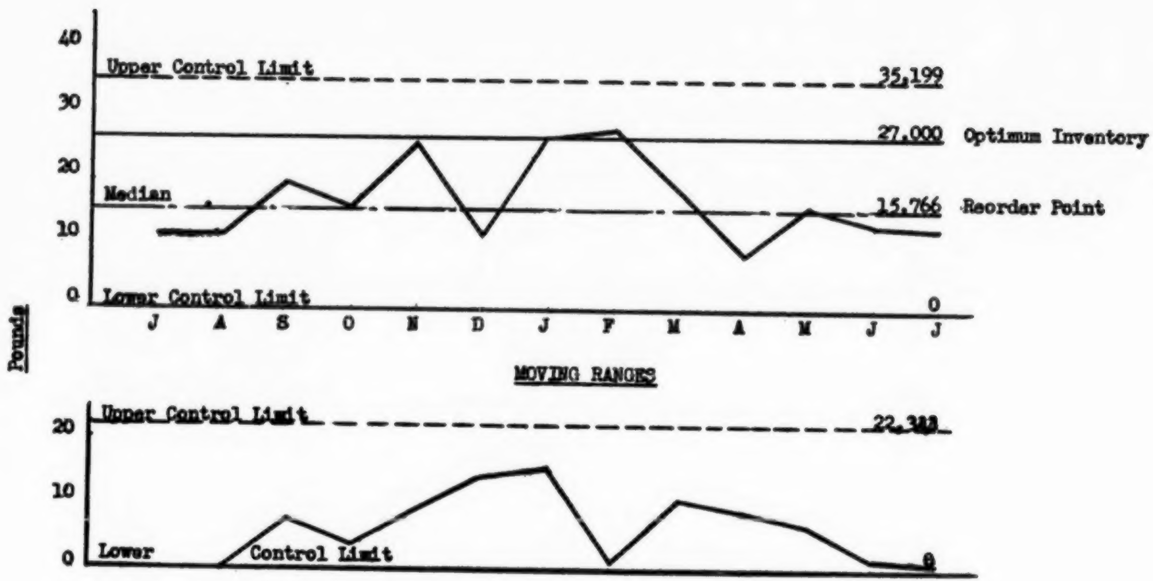


Figure 1- Control Charts for Inventory Control

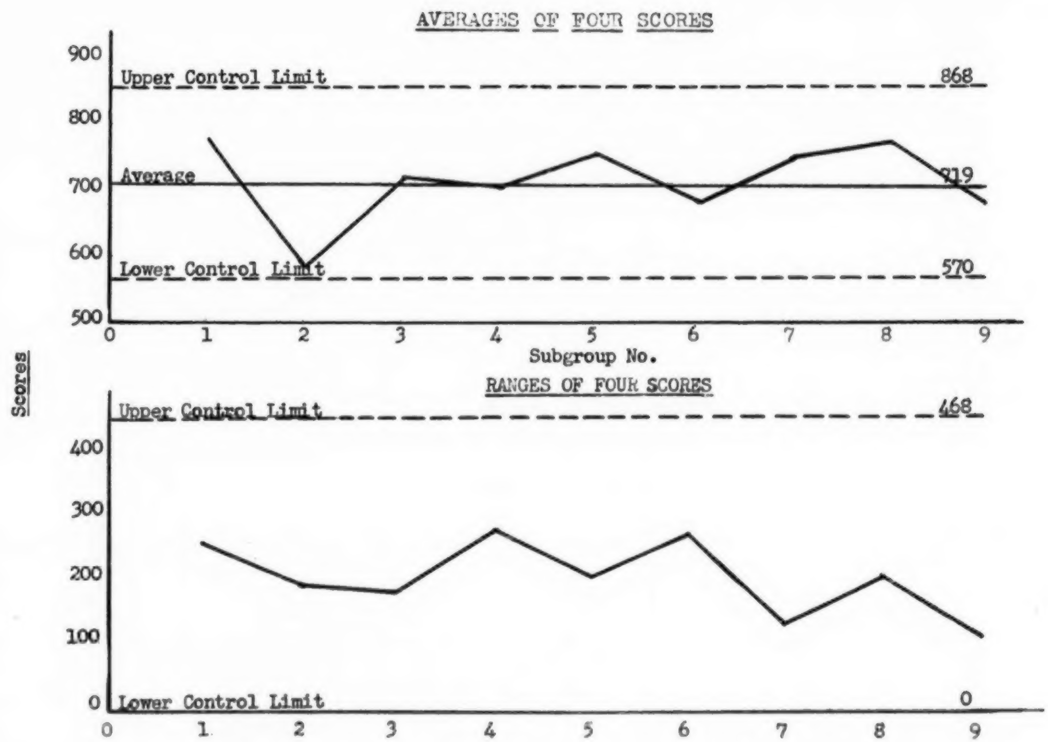


Figure 2- Control Charts for Group N Merit Ratings

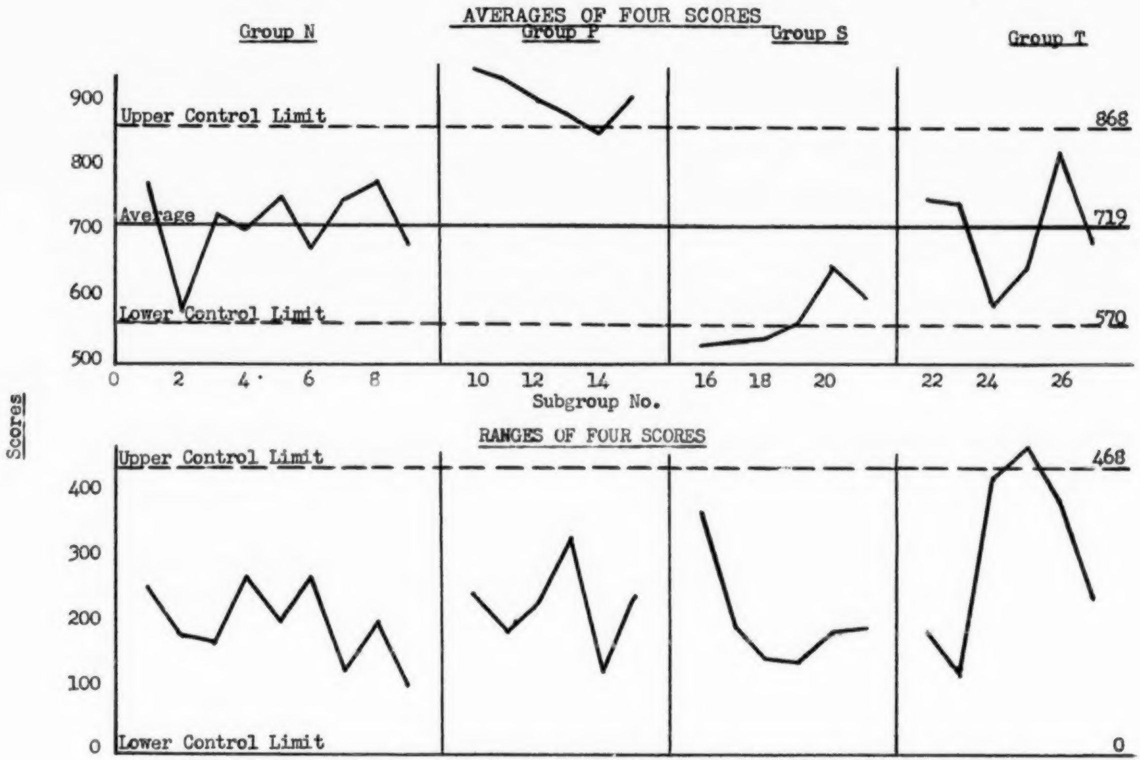


Figure 3- Control Charts for Comparison of Group Merit Ratings

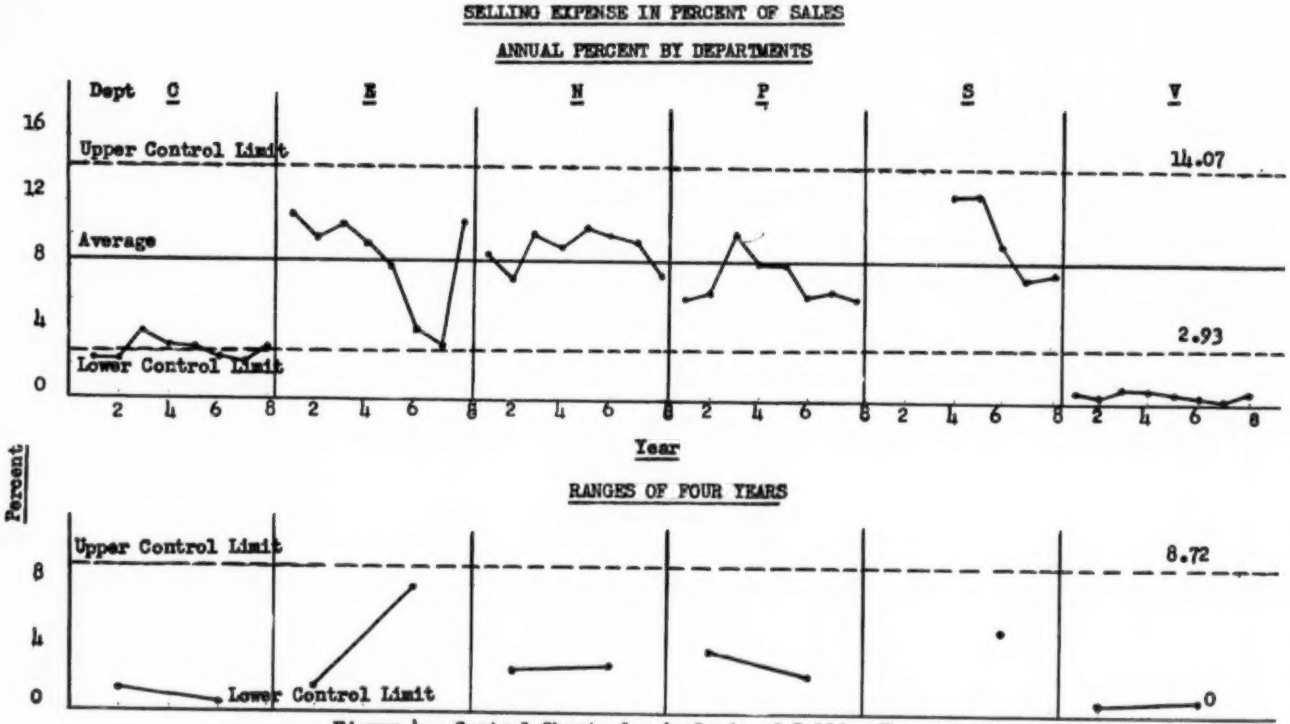


Figure 4 - Control Charts for Analysis of Selling Expense

QUALITY CONTROL TECHNIQUES FOR ESTABLISHING INDUSTRIAL STANDARDS

Ralph E. Wareham
Consultant on Quality Control

The past year has once more brought the return of closer balance between the supply of goods available and the demand for these goods. Periods of national emergency, such as the Korean conflict, create demands for goods far in excess of the available supply. The pressures attendant to these demands of consumers for immediate delivery of goods always has a direct effect on the quality standards maintained. This pressure gradually forces down industrial standards of quality. Once these standards have been lowered, a substantial period of time and much special effort are required before the original quality levels can be regained.

Much progress has been made in the past two years in regaining the desired product quality levels. However, the practical situation at this time is that problems of quality standards still rank high among the unsolved industrial quality control situations. Those companies which have completely solved such problems are fortunate indeed.

Definition

By way of initial definition, industrial quality standards are defined as comprising industry product standards, commercial guarantees, as well as specifications and requirements agreed upon by the manufacturer and purchaser.

Most industry associations have taken steps toward setting product standards and these steps usually include basic specifications and tolerances. In addition to these industry product standards, a manufacturer may establish certain commercial guarantees for his product; these frequently require closer control than permitted under the industry product standards.

Finally, the specifications and the requirements established by agreement between the manufacturer and customer must be rigidly adhered to. Thus the industrial quality standard for any product is determined by actions taken in the industry as well as by the individual company.

Conformance to Required Quality

The quality problem of greatest importance in most companies is delivering products which fully meet all customer requirements. In many cases this has proved difficult--not because of poor process capability or because of lack of effort to meet customer requirements--but rather on account of differences in interpretation of quality standards between the two or more companies involved. Many here in attendance today have made field trips to customer plants on complaint calls only to find that the key problem was one of interpreting the quality standard required. In some cases, the customer's quality standards may be much stricter than the supplier's previous quality limits. In other cases, the customer may even be rejecting product for a characteristic not covered by your own inspection and testing. In still other caees, you may have found that the characteristic given greatest emphasis in your own inspection may be counted of little

importance to the customer.

Therefore, an important part of the quality standards problem is one of communications. Frequently such communications cannot be handled by correspondence or by telephone conversations; a closer contact between the quality control groups of the customer and the supplier is needed. Most companies have found that their quality control people must make field trips to customer plants so as to secure first-hand information on quality standards required and acceptance procedures to be followed.

One company with a very successful quality control program has established the procedure of having a quality control supervisor present during the first delivery on any new contract. This plan is, in part, possible due to services of company-operated aircraft, which can reach any customer's plant within a few hourse. While this special service involves substantial additional cost, the benefits in reduced complaint expenditures have been more than ample to cover the costs involved.

Techniques Required

Complete quality standards for any product require that both visual and measurable characteristics be accurately defined. The relative importance of visual vs. measurable characteristics will, of course, vary for different products. However, both must have quality standards which can be accurately interpreted by both manufacturer and customer and by their inspection and test organizations.

The examination of product for visual characteristics requires classification of individual units as conforming or not-conforming with the specifications. These specifications may require that characteristics such as color match, surface uniformity, and satisfactory general appearance be held to close limits. Therefore, techniques are needed for determining which units of product meet and which do not meet the desired visual standard.

Measurable characteristics are usually closely defined with numerical limits in the applicable specifications. However, differences frequently in interpreting the specification requirements. A meeting-of-minds is needed between the manufacturer and customer as to what constitutes compliance with the specification. For many industrial tests new quality control techniques are being developed for this purpose.

Quality Standards for Attributes Testing and Inspection

In attributes-type inspection, where the product must be visually inspected or otherwise compared with product standards, the problems of interpreting quality requirements are especially important. For the most part, this is due to the difficulty in maintaining agreement on the quality standards between the companies involved. The problem of uniform interpretation is also present within inspection and testing groups of a single company and must be constantly checked and kept under observation.

Many companies have found that, without definite reference points, quality standards tend to shift from time to time--often drifting far from original levels. Furthermore, trouble encountered with one type

of defect may cause special attention to be given this one defect, while overlooking others.

Agreement must, therefore, be reached between manufacturer and consumer as to what constitutes a defect and the classification of these defects as to severity. Then standards must be set for reference purposes. Finally, instructions concerning the defect involved must be communicated to all persons involved in the inspection and testing operations, both in the manufacturer's plant and in the consumer's organization.

During the past five years, much progress has been made in handling these attributes inspection problems by scientific means.

Quality Control Techniques for Attributes Standards

We may now inquire as to what quality control techniques are of value in establishing attributes standards. A survey of industrial products shows that many such techniques have been used to advantage. However, the following techniques have wide applicability:

1. Defect Classifications: Extensive use has been made in applying classifications of defects in attributes inspection problems. This has been due, in part, to military requirements for such classifications of defects under government contracts.

In MIL-STD-105A covering "Sampling Procedures and Tables for Inspection by Attributes"[1], a classification of defects is defined as being an "enumeration of possible defects of the unit of product classified as to their importance." This same principle of classification has served industry well in focusing attention on the most important defects and in better evaluation of over-all product quality.

2. Approved Samples: The exchange of approved samples between supplier and customer at the start of a new contract has proved very worthwhile, particularly where close appearance standards must be maintained. Such approved samples provide reference points for the manufacturer both in his manufacturing and in his inspection. They also help maintain uniformity of judgment between the inspection departments of the two companies.

Approved samples also provide means for controlling quality on subsequent product runs for customer reorders.

3. Reference Samples: Reference samples in the manufacturing and inspection areas of a plant are of high value in providing answers on attributes quality standards as soon as they occur. In many industries, questions regarding individual quality characteristics arise frequently from shift to shiftand day to day. The use of reference samples permits a meeting-of-minds between production and inspection as to what is required. Prompt decisions can thus be made on the manufacturing floor.

In providing reference samples, an important problem has frequently arisen in providing samples which do not rapidly deteriorate with time and use. Here much initiative is required to develop a plan that will work in the individual situation; many companies, however, have succeeded in overcoming this obstacle.

4. Rating Procedures: These rating procedures provide a means for graduating quality beyond a "pass-or-reject" classification. They are designed to grade product quality along a continuous scale, so as to indicate how closely the desired standards are being maintained.

Experience indicates that, when such a rating system is developed, accuracy of inspection is improved. In many such rating plans, repeat checks on the product can be made with good precision.

Attributes rating procedures usually involve considering both the prominence of defect and frequency of the defect. The effects of both prominence and frequency are combined in the quality rating as a number. This rating statistical treatment in analysis of quality results.

5. New Instrumentation for Measurement: The area of testing and inspection which must be performed solely on an attributes basis is steadily narrowing. New instruments for measurement and new techniques for testing are largely responsible. It appears likely that this trend will continue if not accelerate.

Quality control departments must be constantly on the alert for such new methods of measurement which might apply to their attribute inspection problems. Frequently, much work is required, however, before the particular instrumentation can be applied to one's own needs.

Quality Standards for Variables Testing and Inspection

The matter of agreement on quality in variables-type inspection and testing is likewise complicated. The measurements and tests made by the producer may not agree with those made by the customer. This will lead to friction between the two organizations and rejection of individual shipments.

There may be many reasons for such lack of agreement in measurements and tests made by the two organizations. However, when such differences do occur, the situation is indeed baffling. An air of uncertainty is thrown over the entire testing programs of both companies. Much effort must then be expended before the situation is clarified.

In many cases, lack of agreement in test results may be due to differences in the type of test equipment used. Frequently, two instruments designed for the same type of test have different recording scales; in other cases, the unit of measurement may be entirely different. Either case can lead to difficulty of interpreting and comparing test results of one laboratory with those of another.

Another source of difference may be the test procedures themselves, where different test and inspection groups may be using basically different procedures in making the tests. Even where industry test standards have been issued, there is freqnently room for difference in establishing the test procedure.

Finally, even with approved test and measurement equipment and with established test procedures, errors can easily creep into product evaluation due to deviations in test and measurement equipment from

lack of adjustment and calibration. Such deviations are present more frequently than most of us would like to admit. They present special problems in the area of customer-vendor relations.

This area of accurate quality standards for variables inspection is now receiving closer attention in many important industries.

Quality Control Techniques for Variables Standards

The procedures for developing accurate quality standards for measurable quality characteristics have received much attention in industry. This work has been done both by industry associations and by individual companies. The following techniques have proved of general applicability in establishing such variables standards:

1. Test Method Standards: Industry-approved test procedures go far toward removing differences in test results. These standard test methods must be provided in considerable detail, if differences in actual performance of the test are to be avoided.

In cases where frequent difficulties between laboratory test groups are encountered, it has been found desirable to establish check lists on the method of making the test. Then the test supervisors check actual practice against these lists from time to time.

2. Test Equipment Standards: Where the industry can establish test equipment standards, better agreement on test results is obtained. Such equipment standards assist the manufacturers of test instrumentation in developing standardized test scales and equipment operation. This overcomes the problem of interpreting test results between two different scales of measurement.

Where no industry standards exist, the companies directly involved can establish one or more instruments as standard for their testing. This again removes the factor of interpreting different scales of measurement.

3. Calibration Control: Even the best measuring equipment drifts out of calibration from time to time. During these periods, test results mean no more than they would from a crude measurement method; often they give completely wrong answers.

Instrument and gauge checks made on a planned schedule provide an effective means for keeping measurements accurate as far as calibration is concerned. They employ the same principles of scientific sampling as are used for product evaluation. Under Air Force contracts, such gauge checks are required in complying with quality control procedures under MIL-Q-5923B.[(4)]

4. Test Repeatability Control: Where extensive routine testing is required, procedures are needed for accurate control of test repeatability within the single laboratory. In making these repeatability checks, duplicate tests are required on the same material under the same test conditions. From these duplicate tests, statistical measurement of test differences within the laboratory can be accurately determined. This repeatability control follows the same basic sampling principles as used in product evaluation and is interpreted by basic statistical rules.

5. Statistical Analysis of Data: The use of statistical principles in analysis of test data provides more accurate interpretation of test results and often permits reduction in amount of testing required.[3]

The wide-spread industrial use of statistical methods in handling test data has brought agreement in the method of reporting quality and has provided a means for checking results between laboratories. This has led to better interpretation of specifications and tolerances.

Understanding of statistical control methods and probability principles has enabled industry to recognize that absolute agreement in test results is not to be expected, but that control within definite limits is the objective.

However, the most important result of statistical control procedures has been establishment of a scientific method for determining true process capability. This has enabled industry to set proper specifications for standard values and allowable limits. The benefits of such precise specifications are now being realized and will prove of even greater value in years to come.

Summary

The quality control techniques discussed in this paper have served industry well in establishment of closer quality standards and in maintaining these standards. They have ben a part of the great advance in basic quality and in quality uniformity that has come since the end of Worl War II.

References:

(1) MIL-STD-105A, "Sampling Procedures and Tables for Inspection By Attributes", Superintendent of Documents, Washington, D.C., 1950

(2) MIL-Q-5923B, "Quality Control of Aircraft and Associated Equipment", U. S. Air Force, Air Material Command, Dayton, Ohio.

(3) A.S.T.M. Manual on Quality Control of Materials, American Society for Testing Materials, Philadelphia, Pennsylvania, 1951.

THE QUALITY CONTROL PROGRAM OF THE DEPARTMENT OF DEFENSE

Brigadier General William P. Farnsworth, USAF
Office of the Assistant Secretary of Defense (Supply and Logistics)

Last fall, when your Program Chairman invited me to present a paper at this Convention, I was delighted to accept because I anticipated that by May I would be in a position to bring you particularly good news regarding the progress of Quality Control in the Department of Defense. I am happy to report that Quality Control is now recognized as a top level management function of the Department, with responsibility for broad policy direction assigned to the Staff Director for Inspection and Quality Control of the Office of the Assistant Secretary of Defense (Supply and Logistics).

As you no doubt know, the Office of the Secretary of Defense, of which Inspection and Quality Control is a part, is a comparatively small, non-operational organization that establishes broad policy for the guidance of the Department of Defense as a whole. The Department of Defense is, of course, a mammoth organization with a multibillion dollar budget. As part of its military activities, the Department provides logistic support for our far-flung Army, Navy and Air Force. The term "logistic support" covers a wide area of activities including the determination of requirements for, as well as the procurement, maintenance, storage and transportation of supplies. These activities and the manner in which they are managed have immense repercussions in industry and, in fact, on our total national economy. It is, therefore, particularly gratifying that an organization of such magnitude as the Department of Defense recognizes Quality Control as an essential element of top level management and logistic planning.

My purpose in talking to you is threefold: first, to outline the fundamental Quality Control philosophy of the Department of Defense; second, to indicate how this philosophy is being implemented; and third, to suggest how industry can accelerate the progress of the Department of Defense Quality Control Program. In this latter respect, I am particularly concerned with soliciting industry's cooperation not only in the interest of promoting better quality materiel, which is reason enough, but also because I am convinced that on a strictly profit basis industry can promote its own interests while reducing costs to the taxpayer.

With respect to the basic philosophy of the Department, I feel that I cannot improve upon what the Assistant Secretary of Defense (Supply and Logistics), the Honorable Thomas P. Pike, has said on this subject. With his permission, I shall quote from a memorandum he wrote quite recently to the Acting Director for Cataloging, Standardization and Inspection, Mr. Nathan Brodsky, expressing his considered views on the scope and function of Quality Control. I think that Mr. Pike's memorandum reflects the kind of deep insight and broad perspective that is especially warming to us, as members of the American Society for Quality Control, who feel that with proper management support and understanding Quality Control has a vast unrealized potential for contributing to industrial efficiency. Let me, then, quote what Mr. Pike has to say:

> "The increasing demand for military equipment and supplies of higher performance and greater reliability has focused attention on the scope, objectives and effectiveness of Inspection and Quality Control throughout the Department of Defense.

Because of the vital relationship of quality to the readiness and reliability of military equipment and supplies, it is essential that Department of Defense policies emanating from this office reflect a comprehensive view of the function of Inspection and Quality Control in the overall Supply and Logistics Program.

In this connection, I wish to state some broad principles which are applicable to the administration of the Inspection and Quality Control Program of your Directorate.

(a) Inspection and Quality Control policies must encompass all materiel entering supply channels, regardless of whether such materiel is procured from industrial sources, is fabricated at a government facility or is obtained from maintenance, supply and storage activities.

(b) The maximum benefits of an Inspection and Quality Control Program cannot be realized unless the various facets of the program, and Inspection and Quality Control's relationship with other activities, are properly coordinated at a high management level.

(c) In achieving its primary objective of assuring product quality, Inspection and Quality Control is to be viewed as a constructive activity directed towards the prevention of defects, the detection of unsatisfactory trends, the conservation of material, manpower and equipment, and towards the pooling of meaningful quality data for utilization in design, maintenance and production, and in supply management.

(d) Since the ultimate measure of quality is effectiveness and reliability in service, the accurate and definitive evaluation of product quality necessitates the feed-back of performance quality information to Inspection and Quality Control for appropriate action and for use by other interested activities.

(e) The objectives of Inspection and Quality Control are most effectively achieved in collaboration with, rather than by duplication of, other activities within both Government and industry. With respect to the latter, maximum utilization must be made of contractors' inspection and quality control information.

(f) To be effective, Inspection and Quality Control must be dynamic and, therefore, must incorporate new technologies and new skills in order to keep in synchrony with parallel developments in the fields of design, production, maintenance, and industrial management generally."

Mr. Pike's statements are really the distillation of many hours of discussion and study of Quality Control's relationship to the Department's ultimate objective of winning the battle for peace by strengthening our industrial machine. He has implemented his views by specific directions, not pertinent to this paper, for the development of the Department of Defense Quality Control Program.

For purposes of this talk, I would like to classify under five

headings the motives and background considerations that underlie Mr. Pike's philosophy. Naturally, categorization of any kind tends towards over-simplification, but I think the following five topical headings provide a frame of reference for discussing the Department's basic point of view. These are: (1) the implications of mass production to quality, (2) the complexity of the design of modern military equipment, (3) the scope of the Department of Defense Military Supply System, (4) the economic repercussions of Quality Control, and (5) the need for measuring quality not only in terms of conformance to specifications but also in terms of performance in service.

MASS PRODUCTION -- The subject of mass production in its relationship to Quality Control is so extensive that it could provide material for a good-sized book. Mass production, of course, goes back to the dawn of the Industrial Revolution and since that time its techniques have been in a constant process of evolution towards greater intricacy and acceleration. But, regardless of the present or future technological status of mass production, we should recognize at least two facts with respect to its relationship to Quality Control. First, that a production process, however well-engineered, can sometimes generate defective materiel at the same high rate that it previously produced conforming materiel. Because of this fact, it is the function of Quality Control not only to detect defectiveness but also to give an alarm as soon as possible that all is not well. It is not enough for Quality Control to function as a police activity concerned only with an after-the-fact separation of defectives from non-defectives. Second, mass produced items, particularly those related to military materiel, are frequently components of larger assemblies, not end-products in themselves. These components must be replaceable; replaceability implies the need for interchangeability. It is only when variations in quality are kept to a minimum that interchangeability can be achieved. Inventiveness and skill of a high order are necessary to minimize quality variations, particularly in the precision industries. In the final analysis, the control of quality is in the hands of the producer. It is my own thinking that for complex military equipment the Department of Defense has a right to require contractors to maintain appropriate controls. Incidentally, by "appropriate" I mean various kinds of controls, not necessarily those exclusively of a statistical nature.

COMPLEXITY OF DESIGN -- Despite the extensive resources of personnel and facilities that are available to the Department of Defense, it is very often technologically and economically unfeasible for the Department to inspect and test an end-item to the degree that conclusive assurance of operability is attained. One reason for this, among others, is that much of our military equipment, such as airplanes, guided missiles, tanks and fire control equipment, are vastly complex in design and cannot be tested conclusively for such characteristics as reliability and life. We also know that many components of a final assembly may be inaccessible for testing in its assembled configuration, or certain critical characteristics may be of such a nature that they cannot be tested except by test to destruction. Sometimes gaging and instrumentation becomes so intricate and so expensive as not to warrant the time and money necessary for Government testing and inspection. I mention these considerations only to emphasize the need for projecting our Quality Control thinking so that we can formulate a "modus operandi" by which maximum and objective quality assurance is achieved during manufacturing, and also by which the consumer and the vendor can make joint use of quality control data. The complexity of our equipment is a salient factor in forcing us to abandon a "caveat emptor" philosophy, and to recognize the technological fact

that our military equipment must be made right in the first place and that we must know that it is.

ECONOMIC REPERCUSSIONS -- The economic repercussions of Inspection and Quality Control are more far reaching than they appear on first sight. These repercussions are of two types. The first relates to the conservation of our total resources including man-hours, machine-hours and raw materials. The second relates to competition among suppliers for contracts with the military departments. With respect to the first, it is quite obvious that Quality Control cannot be considered as something divorced from the economic environment within which it operates. Products that approach perfection can conceivably be manufactured, but their cost would soon drive us into national bankruptcy. Actually, what we need is a practical balance between quality, to satisfy our needs, and the ability of industry to produce that quality within reasonable limits of overall cost. Quality, then, must be related to objectives. It is not always necessary to purchase the best quality possible. The economic relationship between Quality Control and national defense must be considered in terms of cost, capacity for production and realistic quality standards.

With respect to the second consideration I mentioned above, namely, competition, it is well to keep in mind that Quality Control plays a key role in maintaining the stability and integrity of the Department of Defense competitive procurement system. Contractors have a right not only to bid for contracts, but also to have their product evaluated objectively by the government. Such objective quality evaluation assures that one producer does not have an economic advantage over another.

Keeping these thoughts in mind, it is quite apparent that, for economic reasons, the Quality Control policies of the Department of Defense must be so formulated as to conserve resources, to encourage the prevention of defectiveness, and to assure that all producers receive a square deal in their relationship with the government.

SCOPE OF THE MILITARY SUPPLY SYSTEM -- The supply system of the Department of Defense is so mammoth that it discourages description. At the risk of oversimplifying the situation, we might think of supply as a single main pipeline with three major feeder lines, namely, procurement, storage, and maintenance and overhaul. When, figuratively speaking, a soldier, sailor, or airman opens a valve on the main pipeline, ammunition, guns, clothing and food flow out. But where do these things come from? The government procures most of them directly from commercial sources and either delivers them directly to the using services or holds them in reserve. Other supplies come from overhaul and maintenance depots. You might think of these latter items as "secondhand" but, in so far as the using services are concerned, an overhauled airplane or overhauled ammunition serves exactly the same purpose as the brand new product. Still other items are drawn from reserves, from storage. These reserves -- our military stockpile -- are stored in warehouses, depots and ammunition dumps throughout the world. It is immaterial to the soldier, sailor or airman whether his supplies come directly from the producer, from a maintenance depot or from a storage bin. His only concern is to be assured that the item will do its intended job. It is evident that it would be folly for the Department of Defense to ignore quality-wise any one of these feeder lines be it procurement, storage, or maintenance and overhaul. The supply system has to be considered an integrated structure like the water supply system of a big city. Once the

main pipeline has been adulterated, the source of adulteration is immaterial to the user.

PERFORMANCE QUALITY EVALUATION -- There was a time in history when it was quite easy to determine the performance quality of military supplies, and it was also reasonably easy to initiate corrective action. When military equipment and the supply system was less complex and our communication system less extensive, information could be collected and fed back to manufacturers without any highly formalized system of data collection, transfer and analysis. Also, in times past, when equipment failed during military operations, there was usually the possibility of a "second chance". In today's world we are less likely to have that second chance. Either a weapon does its job or the consequences of failure may preclude a second try. Since performance quality is the ultimate measure of the success or failure of a quality control program, it is axiomatic that this quality must be determined so that corrective or preventative action is taken at the sources of trouble. There is, incidentally, an encouraging aspect to this performance evaluation problem. As a result of major developments in the field of data mechanization, we now have means for getting information and feeding it back to places where it can do some good. It is encouraging to know that many aspects of performance quality can be promptly measured and reported, and improvements initiated. This feed-back of data is an essential component of a sound quality control program.

I think I have sketched adequately the fundamental philosophy, background and thinking of the Department of Defense, and I should now like to indicate how this philosophy is being implemented. You will recall that I have said that our supply system consists of one main pipeline with three feeder lines. I think it would be appropriate to discuss implementation of policy in terms of each of these lines, namely, procurement, supply and storage, and maintenance and overhaul. In each, the problem of implementing basic philosophy reduces itself to three elements, namely: (1) establishment of uniform policy, (2) implementation of policy, and (3) development of supporting techniques.

PROCUREMENT QUALITY CONTROL -- The first purpose of the procurement phase of the Department of Defense Quality Control Program is to assure that products accepted by the government conform to contractual requirements. The problem, then, is to find ways and means, on both a policy and operational level, by which this assurance can be obtained as economically as possible, with due regard to the technological considerations that I have already discussed. Actually, the Department of Defense has already published a basic policy statement which reads, in part, as follows:

> "Determination of conformance of the product to contract requirements shall be made on the basis of objective evidence of quality and quantity. The Government inspector shall make optimum use of quality data generated by contractors in determining the acceptability of supplies. . ." (Department of Defense Instruction 4155.6, "Department of Defense Quality Assurance Concept and Policy", 14 April 1954)

The manner of application of this policy must be tempered to suit each of the wide variety of items purchased by the Department of Defense. It would, for example, have a different application, at least in degree, in the field of guided missiles or aircraft engines than with respect to

office supplies.

The above policy emphasizes that acceptance should be based on objective quality evidence. This evidence should be of a kind that is normally generated by a manufacturer during his production operations as a necessary element of good production engineering. Complex equipment, of course, requires an extensive set of controls. If a contractor making such equipment establishes that his product is manufactured under controlled conditions that collectively constitute a satisfactory quality control system, the government would need only to evaluate and verify the substantiating facts. This would make it possible to reduce government inspection to a minimum. More important, it would encourage manufacturers to establish process controls designed to prevent defects and simultaneously provide a factual basis for product acceptance. But it seems reasonable that the government should define "a satisfactory system" and should also establish some standard procedures to guide government inspectors in evaluating the effectiveness of a contractor's quality control system. In line with this thinking, the Department of Defense does plan to publish a Department of Defense Quality Control Specification that identifies the essential elements of a satisfactory contractor's quality control system. The Department is also planning to prepare a standard guide to inform government inspectors how to evaluate and verify the system established contractually in accordance with the specification. I should like to stress that both of these publications must necessarily be "least common denominator" types of documents.

I have been speaking more or less of complex items. Now, let me say a word about such items as small hardware, clothing and similar supplies which can be definitively and conclusively inspected prior to acceptance and to which the proposed specification mentioned above may not be applicable. In accordance with the policy previously quoted, acceptance decisions for these items should take maximum cognizance of contractors' inspection data. In this area, however, there is considerable need for clarifying what constitutes adequate inspection data, and also for resolving some problems regarding the interpretation of specification requirements. The Department of Defense is working towards the resolution of these issues. These problems are too detailed, however, for further discussion at this time.

MAINTENANCE QUALITY CONTROL -- After new materiel has been used over a period of time or has been on the shelf, it becomes worn, damaged or deteriorated to the extent that it requires maintenance or repair. Within the maintenance and overhaul function, Quality Control serves primarily to assure that materiel has been returned to a satisfactory state of usability. When the maintenance or overhaul is accomplished by a government facility, the government is in effect a manufacturer and, because of that fact, Quality Control must play the same role in government maintenance operations as it does in any well-managed industrial activity. Quality Control must serve not only to assure product quality, but also to support productivity by being properly integrated with production. This can be accomplished only through the utilization of modern analytical and engineering techniques. A typical repair facility, either governmental or industrial, is essentially no different from any other manufacturing establishment, except that each unit processed must be treated more or less separately. The repair required on one unit may not be required on the next. But it is still necessary to maintain a quality control system so that all products shipped from the maintenance and overhaul facilities conform to military quality requirements. As in the case

of procurement facilities, it is not enough to sort good products from bad. Prevention of defectiveness is required if the maintenance and overhaul operation is to be conducted economically. Nor is it sufficient merely to repair equipment without taking corrective action to eliminate the causes of unwarranted failures. Thus, there is need for a data feed-back program within the facility and in coordination with design and procurement activities.

The Department of Defense has not yet promulgated a maintenance quality control policy as it has in the field of procurement. I feel, however, that such a policy should incorporate the following thoughts: (1) that Inspection and Quality Control are essential elements of an effective maintenance organization, (2) that Inspection and Quality Control should be organized and directed so as to assure conformance to quality standards at minimum maintenance costs in terms of materials and manpower, and (3) that Inspection and Quality Control should be industrially integrated to protect quality while at the same time contributing to productivity.

In order to translate the thinking of the Department of Defense into action, we will do two things: (1) prepare a Department of Defense Quality Control Manual incorporating the basic elements of an operational, management-directed quality control system, and (2) develop supporting techniques in the fields of administration, engineering and statistics. Fortunately, in these latter areas the Departments of the Army, Navy and Air Force have done outstanding and prolific work. What is now needed is more widespread acceptance and general application of theoretical work and practical models already in existence.

QUALITY CONTROL, AND SUPPLY AND STORAGE -- We have already said that the government should assure itself that supplies and equipment received from producers conform to the quality requirements of the Army, Navy and Air Force. It seems completely logical, then, that we should go one step further and say that once materiel is in the custody of the government, the government should make sure that this materiel continues to conform to original design requirements.

We know that very few things are completely inert; almost everything deteriorates in storage. It is, thus, quite evident that deterioration must be carefully watched. Of course, when dealing with a mammoth storage program, it is an extremely difficult problem to measure deterioration because of its elusive and long-term nature. However, numerous techniques are available for placing storage surveillance on a sound scientific basis. When these techniques are properly knitted together, they make possible a formalized program of modern administrative, statistical and engineering methodology. Many organizations within the Department of Defense have already, with much imagination and inventiveness, developed such programs to a high degree of operational effectiveness.

In supply and storage the Department of Defense intends to develop its program in the sequence of policy, implementation of policy and the development of supporting techniques. The Department has not yet issued a policy statement. For purposes of this paper, I have written some statements that reflect my opinion of Quality Control's role in this area: (1) Inspection and Quality Control are essential functional elements of supply and storage operations throughout the Departments of the Army, Navy and Air Force. (2) Inspection and Quality Control operations should be organized and directed to provide continuing periodic technical

evaluations of the quality status of supplies and equipment in storage. (3) The extent and frequency of such periodic quality evaluations should be based on quantitative and objective analyses of inspection and test data, of environmental conditions and of performance or functional test results. (4) Quality evaluation procedures should incorporate modern technical, engineering, statistical and quality control procedures in order to assure accuracy, reliability and objectivity of quality information.

Until such time as the Department adopts an official view of Quality Control's function in supply and storage, it is hardly advisable to discuss details of implementation. However, this program is rapidly maturing.

My third purpose in presenting this paper is to suggest what industry can do to accelerate Quality Control's progress and, at the same time, serve its own interests. My suggestions might be classified under two headings, namely: (1) management, and (2) operations. With respect to management, the first and foremost need of the moment is top level management's interest in, and comprehension of, the vital relationship of Quality Control to effective industrial operations. Quality Control is, beyond all else, a management function. Second, management must encourage and aggressively push the development of new techniques or, at least, adapt old techniques to new situations. Quality Control must be engineered; it is not a ready-made, ready-to-wear garment. Third, management must make available to Quality Control the same level of technical talent that is assigned to other segments of management and engineering. Without this threefold combination of management interest, creativeness and technical talent, Quality Control remains simply a textbook writer's dream. But dreaming is luxury that we can't afford in times like these when the international situation demands that we must have military equipment of maximum reliability and readiness. The problem of reliability is so great and so serious that management does a great disservice to itself and to the Department of Defense when it fails to bring to Quality Control some of the abundant skills and driving energies that are characteristic of American industry generally.

Finally, I should like to ask management to view Quality Control in broad perspective, to recognize that the tools of Quality Control are many and varied. The time has come to abandon the antiquated idea that Quality Control is primarily a statistical gimmick. As long as we take a myopic and parochial view of Quality Control, with our attention glued on operating characteristic curves and sigma limits, there is little hope for realizing the broader and more rewarding potentialities that the future has to offer.

With respect to operations, I have two basic recommendations which I make at the risk of sounding platitudinous, but platitudinous or not, they bear repetition: (1) that Inspection and Quality Control operations be planned with the same meticulousness as one would design a new product. Preferably, this planning should be done in advance of actual production operations but, of course, subject to such adjustments and improvements as the production process requires. (2) that Inspection and Quality Control operations be placed in closest possible proximity to production activities. The closer you get to the machines in time and space the better. (3) that adequate records be maintained. This does not mean you have to maintain a paper mill. Records are an essential element of scientific planning and administration. Good judgment

dictates that when observations and measurements are made, they should be recorded as a source of guidance for subsequent action to correct deficiencies. When judiciously planned, these records more than pay for themselves.

The problem of proper records is of particular importance because the Department of Defense is committed to the policy of making maximum use of contractors' objective evidence of quality. Obviously, quality evidence must be recorded if it is to be evaluated and verified. The Department of Defense can hardly be expected to accept a product on hearsay. At the same time, the Department of Defense expects evidence to be meaningful, not recorded merely for "front" or to satisfy what might be considered the whims and fancies of government inspectors.

I have outlined the philosophy and plans of the Department of Defense and have mentioned briefly what industry can do to accelerate the progress of the Department's program. I am fully cognizant of the vast amount of work to be done. This work is of such a nature and of such a magnitude that it can be accomplished only by the joint and harmonious efforts of private industry, government and technical organizations such as the American Society for Quality Control. By discussing each other's objectives and problems at technical meetings of this kind, we strengthen impregnably the bonds that unite all of us in the cause of the defense of our country.

APPLICATION OF COMPUTING MACHINES TO THE SOLUTION OF STATISTICAL PROBLEMS OF AN ENGINEERING NATURE

William E. Andrus, Jr.
International Business Machines Corporation

Statistics has taken a prominent place in the Engineering Research and Development Laboratories in recent years. Engineers utilize more and more the services of the trained statistician both in the organization of experiments and in the analysis of the resulting data. This is where the computing machine, the subject of this paper, plays its part.

IBM MACHINES IN STATISTICAL WORK

In 1952, IBM formed the Scientific Computation Laboratory as an integral part of its Endicott Engineering Laboratories. This Computation Laboratory is staffed by engineers, mathematicians, physicists and statisticians who serve as consultants in mathematical analysis, and act as a computing group to provide the solution to a large variety of engineering problems.

The statistical section of this group, which has the responsibility of handling all problems of a statistical nature, works closely with various engineering departments on projects that are contemplated or are underway. Their services range from the statistical design of experiments, through data reduction, and analysis of test data. In the analysis of test data they use computing machines of varying speeds and capacities to analyze statistical data in a minimum of time, and with little organizational effort.

Communication between statisticians and computing machines is called a "program", which is a method of telling the machine what to do with data being fed into it (input data). The program to be discussed consists of a deck of punched cards which tells an IBM Card Programmed Calculator, known as the C. P. C. (Figure 1), how to operate. These cards insert input data into the machine; they program the operations to be performed; and they program the results such as storing, punching, printing, or any combination of these three operations.

The C. P. C. has some 597 digits of storage available and operates at 100 or 150 operations per minute. An operation may be one of the elementary calculations -- addition, subtraction, multiplication, and division -- or it may include a combination of several of these elementary operations. For example, square roots, logarithms, and tirgonometric functions, which can be computed by iterative procedures, are considered as single operations completed at electronic speed.

TWO GENERAL COMPUTING CATEGORIES

General statistical problems have been broken into two general categories: (1) Data Reduction and (2) Analysis. This differentiation is made because a particular data reduction program may be followed by a number of different analysis programs. An extensive library of both data reduction, and analysis programs has been developed and is maintained on file. This library of standard analysis programs makes calculations available on shorter notice and makes the services of the Computing Laboratory of greater value to Engineers. Consequently, the programs necessary for most statis-

tical problems are already available. The statistician merely selects the required programs and turns the problem over to the operating personnel. A short time later he receives the completed calculations for his evaluation and report.

DATA REDUCTION EXAMPLE

One of the basic data reduction problems encountered quite frequently is the determination of the parameters of a normal frequency distribution. The input data is punched into cards and becomes a permanent, flexible record which may be organized and re-organized as required during the analysis. The input data is recorded in the card either as "raw score" or "raw score and frequency", and the deck of input data cards is read into the C. P. C. together with the desired data reduction program cards. One such program will accumulate as follows:

1. N = Number of raw scores (or sum of the frequencies)
2. Σx = Sum of the raw scores (or Σfx)
3. Σx^2 = Sum of the raw scores squared (or Σfx^2)
4. Σx^3
5. Σx^4
6. $\Sigma (x+1)^4$ for checking purposes

The first three values $N, \Sigma x + \Sigma x^2$ are punched into a card referred to as an output card. Later this card becomes an input card for various analysis programs. From the values Σx^3 and Σx^4, the machine computes α_1, the measure of skewness; and α_2, the measure of kurtosis, and prints the results. It also checks itself by comparing the appropriate sum of the summations with $\Sigma (x+1)^4$. Simply:

$$\Sigma (x+1)^4 = \Sigma x^4 + 4\Sigma x^3 + 6 \Sigma x^2 + 4\Sigma x + N$$

The output card containing N, Σx, and Σx^2 also contains a ten digit identification number of the problem, and a program number to show that it is a result card from a particular program.

This result card, with an analysis program, can now be entered into the calculator, which will compute $\bar{x}$, the mean; s, the standard deviation of the sample; and σ, the estimate of the standard deviation of the population. The program instructs one unit of the Card Program Calculator to print the results; while another unit punches the results into a second output card. The second output card is an input card for other analysis programs, such as:

1. T-Test and Variance Ratio Test: Takes two of these output cards and compares the means and variances for significant differences. (One of the cards may be a population card containing μ and σ.)
2. Bartletts Test: Takes any number of these output cards and tests for the homogeneity of variances.
3. Cumulative Frequency Distribution: Takes one of these cards and computes the ordinate values of the cumulative frequency distribution, or the abscissa values for specific ordinate values.
4. Cumulative Tolerance Distribution: Takes one of these cards and computes the cumulative frequency distribution left and right from some prescribed point within the distribution.
5. Others

These programs are completely compatable with one another both in the form of the input and output data, and in the storage locations of the computed quantities within the machine. This allows elimination of the intermediate steps of punching out result-cards, since by following one compatable program after another, only the desired results need be punched or printed.

ANALYSIS EXAMPLES

Analysis of printing or punching devices are other problems encountered periodically. In the case of the printing device, the engineer wishes to know whether the horizontal print alignment, (how near to a straight line do the characters print), of a printing device under development is significantly better or worse than an existing printing mechanism. The engineer also needs to know what percentage of the characters are out of alignment more than, say .006"; or are there any print positions or characters which show excessive variation. These and other questions can be answered quickly with the proper selection of existing programs that are maintained in the program library of the Scientific Computation Laboratory.

Another problem frequently encountered is that of component analysis. Basic components such as transistors, resistors, and capacitors are manufactured by many companies and have as many different characteristics. For a particular engineering project some of the requirements of these components are quite critical, and require an extensive evaluation of the various types available. This particular problem usually involves large quantities of data, and the above described programs will reduce this data into a manageable and decisive form quickly.

For some components, such as transistors, the engineer may require information on a certain characteristic which cannot be readily measured by the manufacturer. The question arises as to what measure can be recommended that has a direct relationship with the required characteristic, and can effectively give some measure of it. This is a problem in correlation analysis which readily lends itself to machine computation. Problems of this nature require that the data be reduced into Sums, Sums of Squares (or powers), and Sums of Products. The various ramifications of these summations are determined by the analysis we would like to use. The library of programs developed by the Endicott Computing group covers a wide range of these problems from Simple Linear Regression to Multiple Correlation.

REGRESSION AND CORRELATION

As an example involving simple linear regression, consider the problem of determining the differences in life characteristics of mechanical components. These types of components generally have a normal failure pattern but in most cases we are interested only in the rate of initial failures.

Figure 2 is a simplified plot of the cumulative failure distributions of two such components. Note that very few failures occur initially but as the operations build up they begin to occur very rapidly, indicating that the components are wearing out. A straight line fit to this data would be useless for comparative purposes since the distribution function appears to be of the form:

$$y=cb^x$$

where x is the number of operations, and y is the cumulative number of errors. However, this function can be handled by simple linear regression by fitting a straight line to the log function:

$$\log y = x \log b + \log c$$

The input data is punched an (x, y) pair to a card, and when the data deck is run with the appropriate data reduction deck, the C. P. C. will compute log y and accumulate the required summations; N, Σx, Σx^2, $\Sigma \log y$, $\Sigma (\log y)^2$, $\Sigma x \log y$, and again for checking purposes $\Sigma(x + \log y + 1)^2$. When this program is followed by the linear regression analysis program, the machine computes from these summations the following values:

1. r = Coefficient of Correlation
2. $\log b$ = Regression Coefficient (slope)
3. $\log c$ = Intercept
4. σy^2 = Regression Variance
5. σ_b^2 = Variance of the slope
6. σ_c^2 = Variance of the intercept

As in the example, one of the necessary criterions might be that the line must go through zero. Thus, we fit the data to the line:

$$\log y = x \log b'$$

The machine also computes for this line:

7. $\log b'$ = Slope
8. $\sigma_y^{2'}$ = Regression Variance
9. $\sigma_{b_1}^2$ = Variance of the slope

To determine whether this later is an acceptable fit, the machine compares the sum of squares of the line forced through zero with the sum of squares of the least squares line by the Variance Ratio Test, and prints out F, N_1, and N_2 for entry into the table. This test is not conclusive, so the machine also computes t, of Students T Test to determine whether the intercept log c, is significantly different from Zero.

Figure 3 is a plot of the log function forced through zero, and Figure 4 shows the corresponding anti-log functions. The machine prints the various results shown above and also punches certain combinations of them into output cards. One particular combination represents the quantities required for the comparison of slopes. Entering the machine with two of these cards and the appropriate analysis program, it is possible to compare the slopes of two lines as illustrated in Figure 3, and to determine whether the apparent difference in component life is significant or not. The example used here could have been solved by other statistical methods. However, it demonstrates the many variations that can be obtained very rapidly with a computing machine.

Other programs in the Regression and Correlation series involve fitting data to other exponential forms, second and third order ploynomials, and

simple multiple correlation problems. Larger problems in multiple regression and correlation analysis are solved on the IBM Type 701 Electronic Data Processing Machine (Figure 5), located in New York, where a general program has been written which will handle up to fifty (50) independent variables, and to 1022 5 decimal digit observations of each variable. This program will compute, in the maximum case, the inverse correlation matrix, 48 partial correlations, 50 sets of linear regression coefficients, 50 multiple regression coefficients, and 50 standard errors of estimate.

NEW MACHINES AND GREATER SPEED

At the present time, most of the programs for the C. P. C. are being streamlined and re-programmed for the IBM Type 650 Magnetic Drum Electronic Data Processing machine. (Figure 6.) This is a stored-program machine which computes at an average of one hundred and ninety operations per second, and has a magnetic drum with a storage capacity of 1000 or 2000 10 decimal digit numbers. It is expected that the use of this machine will result in a tremendous saving in computing time over the C. P. C., particularly in the data reduction programs. Present indications are that these programs will run 5 to 10 times faster. As an example, a linear regression problem involving 100 observations requires approximately 18 minutes on the C. P. C. for the data reduction and complete analysis, including the forcing of the regression line through zero. This compares with a little over 3 minutes for the same problem solved on the Type 650.

One of the first statistical programs developed for the Type 650 is one which will solve problems in factor analysis by the Analysis of Variance technique. This program is general in that it is limited only by the storage capacity available, and will solve most problems providing the number of digits does not exceed 17,000. As an example, a five factor analysis involving five variables at levels respectively of 3, 4, 5, 6, and 7 contains 2,520 observations. If these observations are four digit numbers we have 10,080 digits which is well within the capacity of the program. The machine will compute all of the necessary sums of squares by summing over the variables 1, 2, 3, 4, and 5 at a time and will punch out these values and the corresponding values of n. An additional program, which will be added shortly, will take the sums of squares, convert them to mean squares and test each with the residual to determine which of the main effects and interactions are significant. A typical problem of this nature is the one that is illustrated in Figure 7. Here we refer to the "bounce" of card reading brushes and how it is effected by the various levels of six variables. The computing required for problems of this type are in general quite cumbersome, but a statistician who has programs such as these available merely defines the parameters of the problem, and turns it over to the machine operating personnel. A few hours later he receives the completed analysis which he knows is correct since the machine is self checking, and the program also contains a mathematical check which verifies the results.

MORE CREATIVE ENGINEERING

The computing machine is playing a vital role in science and industry today, and particularly in the IBM Engineering Laboratories. Problems,

which in the past would never have been attempted because of their size, are now being solved as a matter of course, and the results are obtained before they are obsolete. The computing machine eliminates the drudgery of routine calculation and releases scientific and engineering personnel for more creative work.

REFERENCES

Brownlee, K. A. (1948) *Industrial Experimentation*, Chemical Publishing Company, Inc., Brooklyn, New York

Duncan, A. J. (1952) *Quality Control and Industrial Statistics*, Richard D. Irwin, Inc., Chicago, Illinois

Youden, W. J. (1951) *Statistical Methods for Chemists*, John Wiley and Sons, Inc., New York

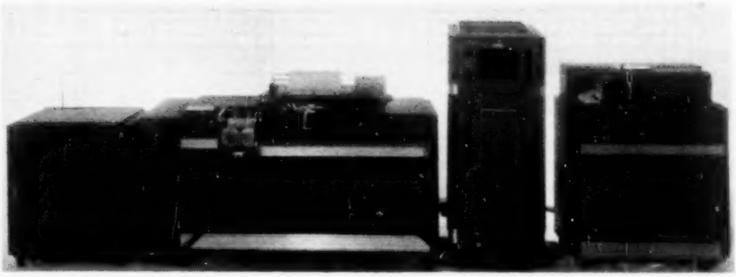


FIG. I CARD PROGRAMMED CALCULATOR

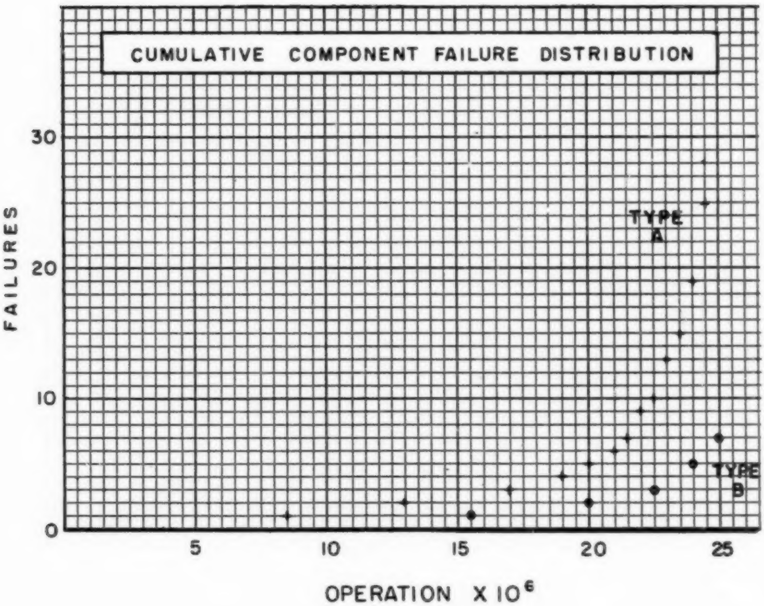


FIG. 2

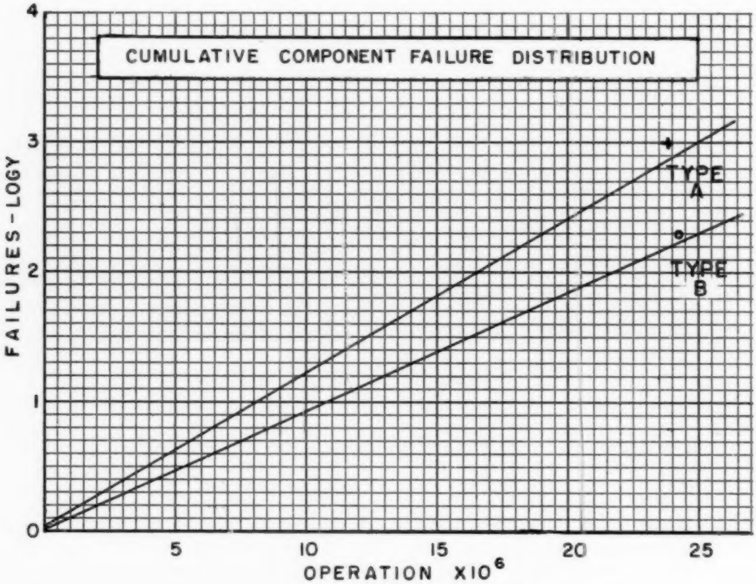

CUMULATIVE COMPONENT FAILURE DISTRIBUTION
FAILURES - LOGY
TYPE A
TYPE B
0
1
2
3
4
5
10
15
20
25
OPERATION X10^6

FIG. 3

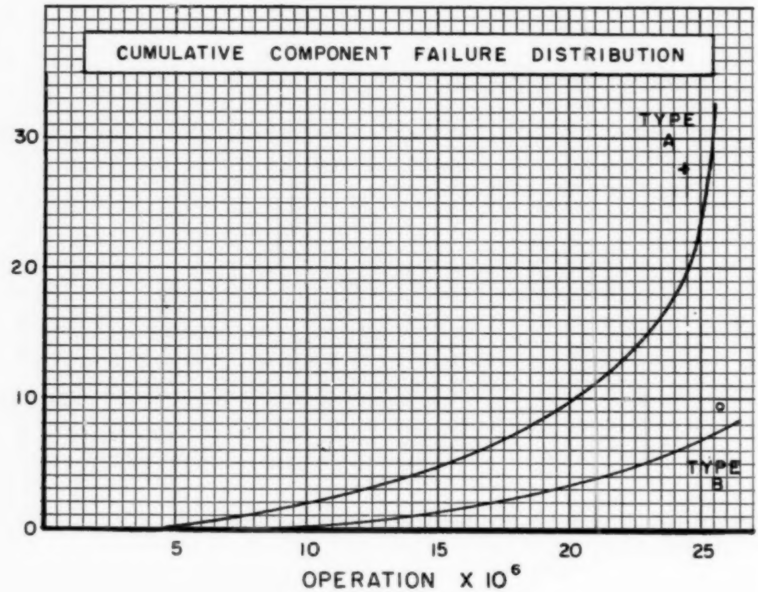

CUMULATIVE COMPONENT FAILURE DISTRIBUTION
TYPE A
TYPE B
0
10
20
30
5
10
15
20
25
OPERATION X 10^6

FIG. 4

FIG. 5 TYPE 701 ELECTRONIC DATA PROCESSING MACHINE

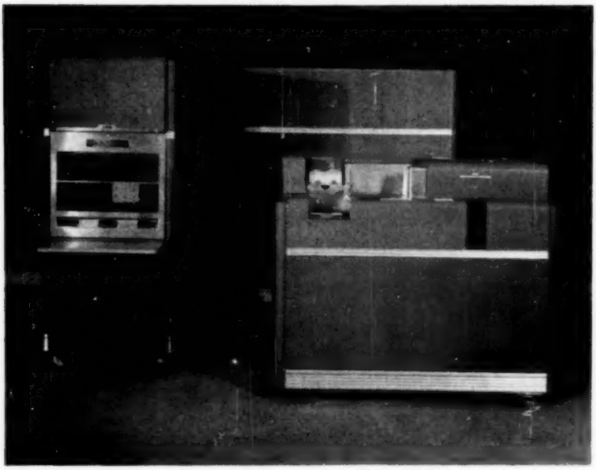
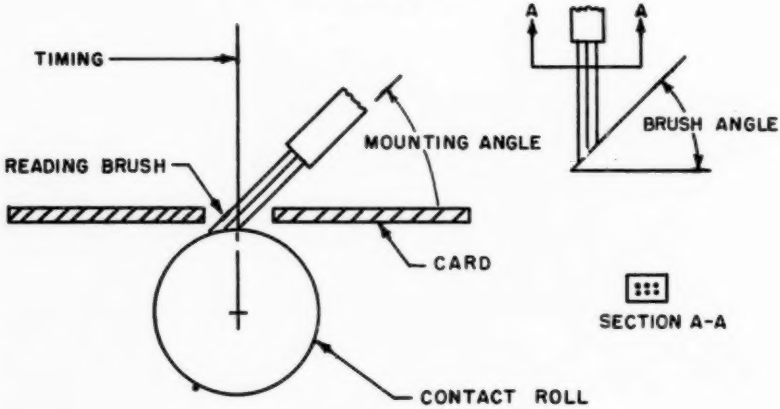


FIG. 6 TYPE 650 MAGNETIC DRUM DATA PROCESSING MACHINE



ANALYSIS OF VARIANCE

DETERMINE WHICH FACTORS HAVE THE MOST EFFECT ON BRUSH BOUNCE

FACTORS ARE:

1. CONTACT ROLL MATERIAL — BRONZE, COPPER, BERYLLIUM
2. BRUSH MATERIAL — STEEL, ALUMINUM, COPPER BERYLLIUM
3. NUMBER OF STRANDS — 6, 8
4. MOUNTING ANGLE — 20°, 30°, 40°
5. BRUSH ANGLE — 20°, 30°, 40°
6. TIMING

FIG. 7

STATISTICAL INVENTORY CONTROL

W. F. Hoehing
Westinghouse Electric Corporation

Every forward looking manager and owner wants his firm to achieve the status of a "growth company." Simply stated, a growth company is one where the dollar put up today earns the biggest return in the shortest time with respect to its competition at large. To achieve this enviable economic status, the company must obtain a maximum return-on-assets and its maximum share of the market. Other than direct means, such as product design, price, sales effort, quality control, cost reduction, of achieving these objectives, inventory in most cases plays the most important role.

The importance of inventory as a factor in this return-on-asset equation is readily accepted if the balance sheets of representative industrial companies are analyzed. From these analyses, it is common to find that inventories range from 10% to 40% of their gross assets. The questions the prudent manager must ask to assure himself that the large investment in inventories is earning its way among alternative capital investments are:

1. Is the inventory sufficient, deployed and controlled in a manner to achieve the maximum volume of sales?

2. Is the inventory too large thus inflicting a restriction on the ability to earn a high rate of return?

3. Is the cost of operating the productive facilities and controlling the inventory low, thus contributing to maximum profit?

Here is the inventory paradox; not too much, not too little and at lowest cost for highest profit. The solution is the optimum inventory investment which can be achieved by making decisions based upon the costs involved. By recognizing the existence of the costs, the first step toward "Management" control of inventories will have been taken.

In actual practice we know that the control of the details of inventory ultimately finds its way to the operating people or into a machine program where the inventory decisions are made as to how much of what to buy or make and when, for each and every item. "How much of what, when" is the literal equation of inventory. If this equation is to be solved on a quantitative basis in accord with a management policy of maximum return on assets, then there must also be a method of communication so that their general decisions can be effectively and uniformly executed. This requires a system that has the means to absorb and the mechanics to solve for many variables of the demand (sales) and cost functions. Here is where mathematical and statistical methods can be used to extend management's quantitative decisions down-the-line and keep the answers consistent with the cost and service objectives. With such a system instead of the operator of say 1,000 stock items making an average of about 70 inventory decisions each day, trying to consider the effect of about 300 cost and demand values, he would work with management approved standards in the form of procedures into which are built the general decision rules.

In this paper, we will show how we have approached this inventory problem by using statistical methods, with the result that the return-on-asset criterion can be achieved. To cover the entire inventory picture would require discussions of work-in-process, miscellaneous and stores inventories. However, the techniques we shall deal with here apply only to the "stores" category. We define stores inventories as those (raw material, supplies, finished parts and finished goods) which are procured and manufactured in larger quantities than required for the immediate future. The reasons for having stores inventories are:

1. To provide customer service, i.e., to have goods available in the time and quantity required to obtain sales and meet market objectives.

2. To obtain operating economy, i.e., to minimize such costs as procurement, machine setup and transportation.

To accomplish these objectives requires the answers to "when-to-order" and "how-much-to-order" each time. Before explaining the techniques which we use to answer the questions, it is necessary to study the characteristics of a stores inventory account. Figure 1 illustrates the behavior of a typical stores item.

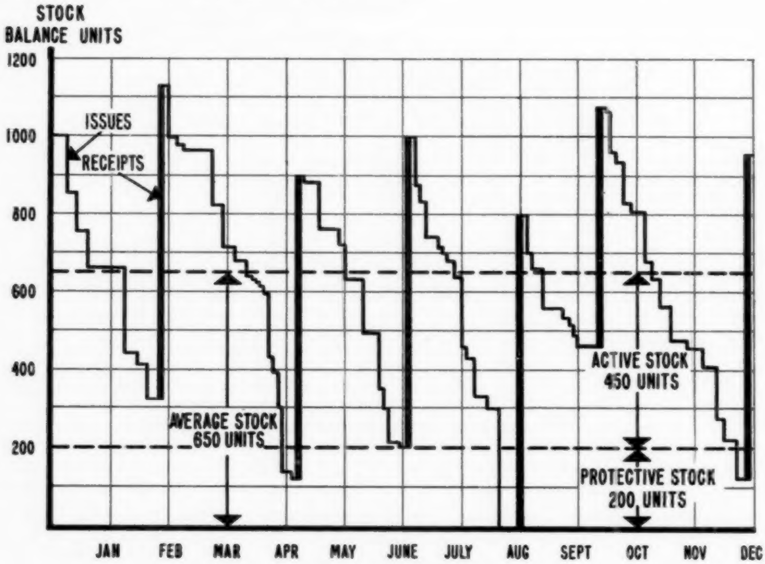


FIG. 1 STORES INVENTORY MODEL

Analyzing the graph we note that there are two distinct parts which we call the "active" stock and the "protective" stock. The protective stock is the average quantity on hand when restocking orders are received. The active stock is the difference between the average total stock and the protective stock. Knowledge of this permits us to effectively apply the service and economy control techniques.

We use the "economical ordering quantity" technique to control the active stock, essentially to achieve the economy objective. In a few words, this technique answers the question as to how much of a given volume of requirements (e.g., what amount of the expected annual sales or usage) should be procured and carried in inventory on the average such that the total of the carrying and restocking costs for the active portion of the stock are minimized. Figure 2 illustrates the economical ordering concept.

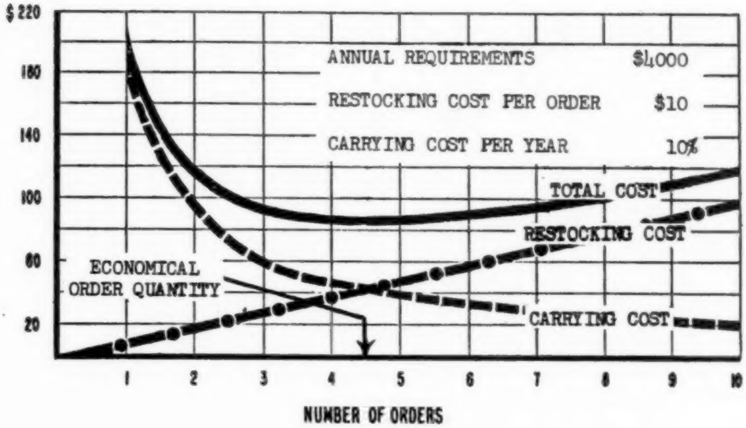


FIG. 2 ECONOMICAL ORDER QUANTITY

Here we note that the costs change with the frequency of the number of orders placed during the year. Actually the total cost is at a minimum where the carrying and ordering costs are equal. This answers the question of "How much to order each time" and is also the point where the active portion of the inventory is at its optimum level.

The solution to the second question, "when-to-order," which controls the degree of service is far more complex and difficult than the answer to "how much." Briefly, stock must be ordered at a point in terms of quantity such that sufficient time is allowed for delivery before the stock is depleted. When to enter a restocking order is controlled by an order point which is based on the expected demand during the delivery time plus the amount of protective stock required to prevent no more than the desired number of stock-outs. Thus, the amount of protective stock and consequently the degree of service can be regulated by increasing or decreasing the order points as illustrated in Figure 3.

From this we can state two premises upon which we can establish the protective stock controls:

1. The only time we can get a stock-out is when a restocking order is open. This is because the order point (if set over zero) will force the

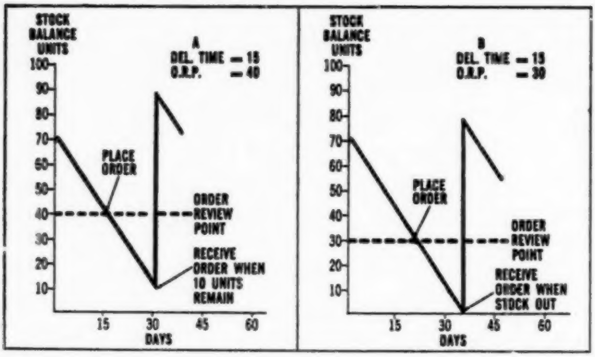


FIG. 3 EFFECT OF ORDER POINT ON PROTECTIVE STOCK, STOCK-OUTS AND SERVICE

placement of an order before the stock can be depleted. Hence, the protective stock problem can be narrowed down to what happens during the restocking time.

2. The frequency of the number of stock-outs is dependent upon the number of restocking orders issued. The number of orders is predicated upon an ordering policy, e.g., the aforementioned economical ordering procedure.

Hence to solve the "when-to-order" and service problems, we must control the inventory in such a manner to limit the number of stock-outs, for example, to one stock-out in a given number of chances. The most desirable condition would be to set order points so that the last unit of stock was used just as the restocking order was received. However, we know that accomplishing this is highly improbable in the long run because it is very likely, when left to chance, that deviations in issues will occur and cause stock-outs. Thus, if we are to set order points with which we can limit the number of stock-outs and consequently control the degree of service, the causes of stock-outs must be determined and allowances made for their occurrence. In reviewing the behavior of stores items, we found that there are three causes of stock-outs -

1. The number of demands (customer or shop orders) can be greater than expected.

2. The size of the demands (units per order) can be greater than expected.

3. The delivery or restocking time can be longer than expected.

To allow for these occurrences and thus control inventories and service, we studied and learned something about their behavior. The problem was solved to our satisfaction by employing probability theory and statistical methods. In the remainder of the paper, we will explain,

in detail, the application of our method to one of the factors that can cause stock-outs, i.e., when the number of demands during the restocking time are greater than expected. After this we will explain the general operation of the procedures which introduce the other two factors, size of demand and delivery time.

In the demand problem we worked with many stock items with the hope of discovering some consistency in their behavior that could be subjected to statistical analysis and control. We gathered data on the frequency of numbers of demands occurring during selected time intervals (per month in this case). The resulting distribution of the number of demands per month (n) for a typical stock item is shown in Figure 4.

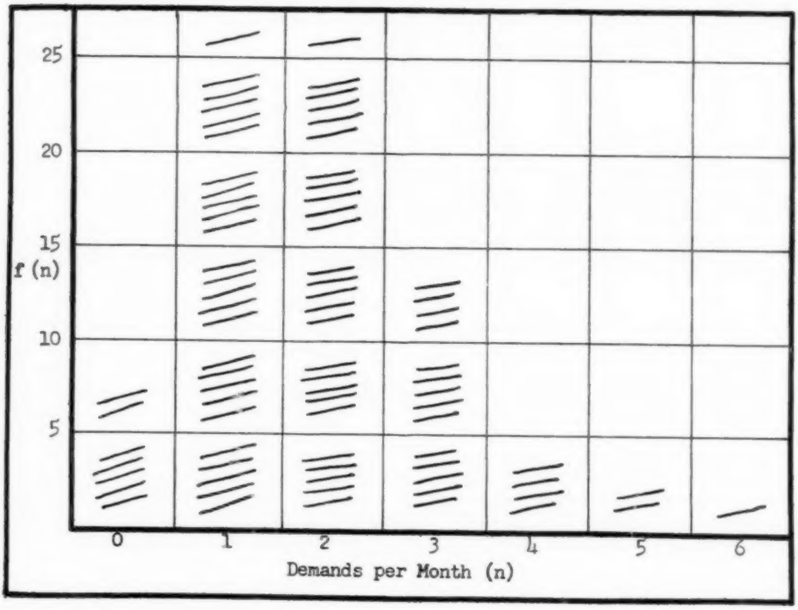


FIG. 4 DISTRIBUTION OF NUMBER OF DEMANDS PER MONTH FOR A TYPICAL STOCK ITEM

It will be observed that the distribution is skewed in a positive direction and that the mean and central tendency lies toward the low numbers. After obtaining similar distributions for many other items because of the appearance we chose to fit and test the data to a Poisson distribution. Figure 5 portrays the results.

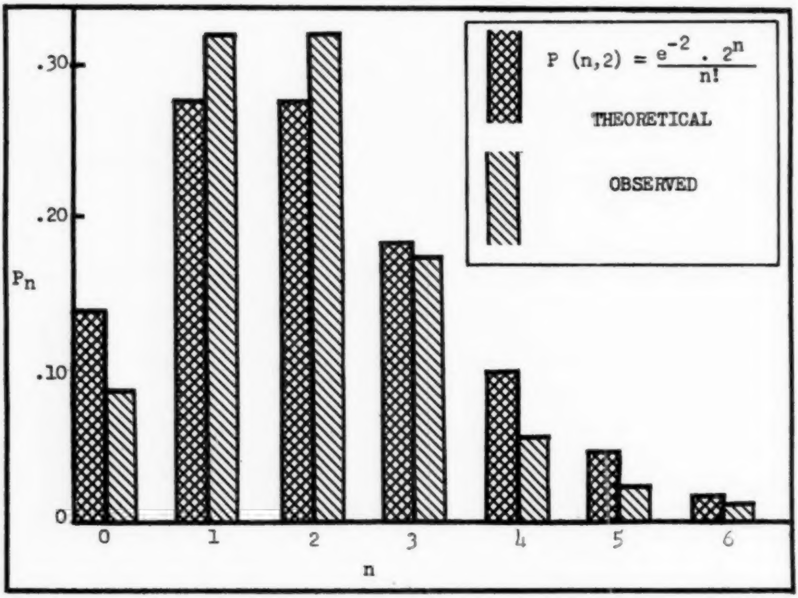


FIG. 5 RELATION OF OBSERVED DISTRIBUTION TO THE THEORETICAL POISSON DISTRIBUTION FOR m = 2

Using the data from the example in Figure 4, where actually 161 demands occurred during 80 months, we calculated the relative frequency which is shown for each value of the number of demands per month (n) up to n = 6. The average number of demands per month (m) in this case was 2. Next, to illustrate the fit, the Poisson distribution for m = 2 was calculated and constructed from the equation:

$$P(n,m) = \frac{e^{-m}\, m^{n}}{n!}$$

After many other studies and tests (χ^2 for the most part), we have concluded that the frequency of demands can be approximated by a Poisson distribution.

The next question is how can this knowledge be put to use in inventory control. Our objective is to provide a given degree of service by placing restocking orders in time so as to control the frequency of stock-outs. Using the same case we can plot the Poisson distribution for m = 2 again as shown in Figure 6.

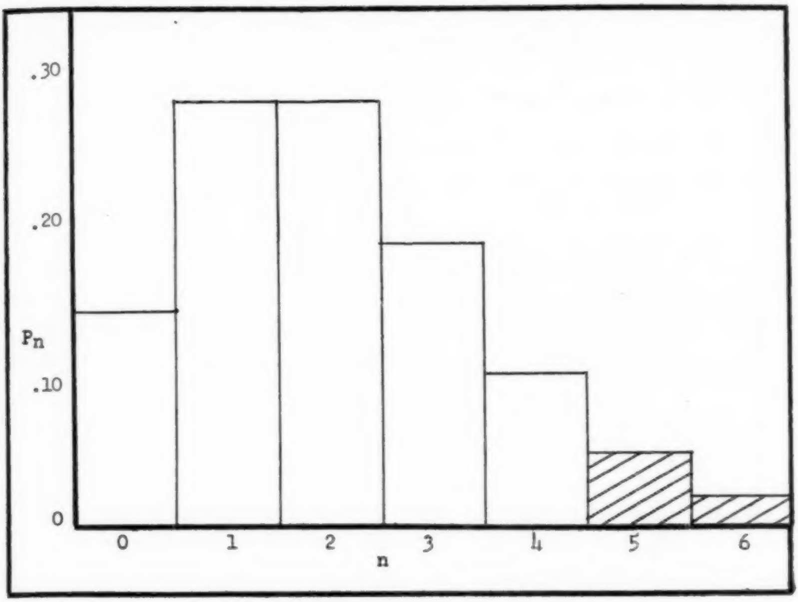


FIG. 6 POISSON DISTRIBUTION FOR m = 2

Then using the distribution for illustrative purposes, we can calculate the chances of a given number of demands being exceeded. (In actual practice, tables similar to the condensed tabulation of Figure 7 would be used.)

n	Frequency	Inverse Accumulation
0	.135	.865
1	.271	.594
2	.271	.323
3	.180	.143
4	.090	.053
5	.036	.017
6	.012	.005
7	.003	.002

FIG. 7 TABLE OF P (n,m) for m = 2

For example, as shown in the inverse accumulation of Figure 7, we would expect that 6 demands would be exceeded 5 times in 1000 chances; 5 demands, 17 in 1000; 4 demands, 53 in 1000; etc. Thus, referring back to Figure 6 we can say that the occurrence of 4 demands per month will be exceeded 5 times in 100.

This procedure gives us a partial answer to the inventory control problem of how much protective stock is required to limit or control the chances of stock-outs occurring. For example, assume that for a stock item -

The restocking time = 1 month.

The average number of demands per month = 2.

The chances of a stock-out occurring is to be limited to 5 times in 100.

This problem is set up in Figure 6, where we note that the shaded area above 4 is about 5% of the total area of the distribution. This would indicate that although we expect only 2 demands during the restocking interval, the time to enter a restocking order would be when the stock balance crossed the equivalent of 4 demands. Hence, we would require 2 demands worth of protective stock to limit the chances of getting stock-outs to .05 or 1 in 20 chances.

The example thus far represents a static condition. To make a practicable application for inventory control, we must be able to handle variable conditions. Hence, we have two further considerations:

A. Perhaps the degree of protection should be varied say from 1 stock-out in 2, 3, ... 10, etc. chances. This means that we must be able to calculate the values represented by an infinite number of areas under the curve.

B. Also, it is known that the average number of demands during the restocking times will vary from one item to another.

To handle the first (A) above, selectivity of protection, we must agree upon a way to express number of stock-outs; e.g.: 1 stock-out per some number of chances which we expect to take during a finite period of time, say 1, 2, 3, 4, etc. years. Thus, we must have a method of measuring the frequency of chances we will take of going out of stock during a period of time. Earlier we established a basic premise which provides the key. It was that we can have stock-outs only when a restocking order is open. It follows, then, that the chances of having stock-outs is a function of the number of orders. Consequently, this can be expressed in the form of a ratio of the number of stock-outs desired to the number of orders placed (chances taken) per period of time (protective period). For example, assume that the protective period chosen was five years and that only one stock-out is desired for an item during that time; also, the number of restocking orders is expected to be at a rate of four per year. Then, twenty orders would be placed in the five year period and only once do we want a stock-out. Therefore, the protective ratio can be stated as 1/20 or .05. By determining the frequency of ordering (which is variable and is calculated from the "economical" or other ordering practice) and establishing the protective period any variable condition can be handled. This now gives us the ability to enter tables of the Poisson distribution (or approximate continuous curves similar to those shown in Figure 8) with a specific protective criterion which, when solved, will give the proper order point in demands.

Before the system can be integrated, however, we must answer the second consideration, B above, that the expected value of demands will vary item to item. This problem can be practicably solved by drawing a set of curves which represent the accumulative frequencies for the expected values of the Poisson distribution. Figure 8 is a set of curves for several values of m.

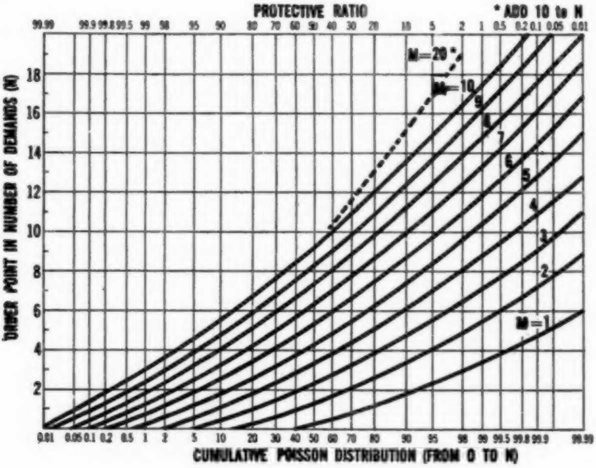


FIG. 8 POISSON DISTRIBUTION-ACCUMULATIVE FOR DETERMINING ORDER POINTS

With the graph, the scheme now can be merged and we can find the order points for various degrees of protection and various expected numbers of demands. The graph can be used to find the order point by calculating the protective ratio and entering the graph at the proper point on the "Protective Ratio" scale (top). Then by proceeding down vertically to the curve representing the expected number of demands, (m), during the restocking or delivery time, the order point in demands can be read, opposite, on the left ordinate (n).

Up to this point we have solved the first problem, that of protection against more than the expected number of demands. The next problem concerns protection against deviations in the size of demands. Realizing that the expected size of demands can be exceeded and also cause stockouts, a certain amount of protection for that occurrence must be considered too. The size of demands is an exponential function which should be calculated independently for a particular stocking location. This value,

ORDER REVIEW POINTS - SMALL NUMBER DEMANDS

PROTECTION: 1 STOCK OUT PER ACCOUNT IN 5 YEARS

AVERAGE NUMBER OF DEMANDS DURING DELIVERY TIME	NUMBER OF TIMES PER YEAR STOCK IS ORDERED																	
	1.0	1.2	1.5	1.7	2.0	2.5	3.0	3.5	4.0	4.5	5	6	7	8	9	10	11	12
.1	.10	.10	.20	.25	.35	.50	.60	.70	.75	.85	.90	1.0	1.1	1.2	1.2	1.3	1.3	1.4
.2	.20	.35	.55	.65	.75	.90	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.7	1.8	1.9	1.9
.4	.65	.80	1.1	1.2	1.3	1.5	1.6	1.8	1.9	2.0	2.1	2.3	2.4	2.5	2.6	2.7	2.8	2.9
.6	1.0	1.2	1.5	1.6	1.8	2.0	2.1	2.3	2.4	2.5	2.6	2.8	3.0	3.0	3.2	3.2	3.3	3.4
.8	1.4	1.6	1.9	2.0	2.2	2.4	2.6	2.8	2.9	3.0	3.1	3.3	3.5	3.6	3.7	3.7	3.8	3.9
1.0	1.7	2.0	2.3	2.4	2.6	2.8	3.1	3.3	3.4	3.6	3.7	3.9	4.1	4.2	4.3	4.4	4.5	4.6
1.5	2.5	2.7	3.1	3.3	3.5	3.7	4.0	4.2	4.4	4.6	4.7	4.9	5.1	5.2	5.3	5.4	5.6	5.7
2.0	3.2	3.5	3.9	4.1	4.4	4.7	4.9	5.2	5.4	5.6	5.7	6.0	6.2	6.3	6.4	6.6	6.7	6.8
2.5	3.8	4.2	4.6	4.8	5.1	5.4	5.7	6.0	6.2	6.4	6.5	6.8	7.0	7.2	7.3	7.5	7.7	7.8
3.0	4.5	4.9	5.4	5.6	5.9	6.2	6.5	6.9	7.1	7.3	7.4	7.8	8.0	8.2	8.4	8.5	8.7	8.8
3.5	5.1	5.5	6.0	6.3	6.6	7.0	7.3	7.7	7.8	8.2	8.3	8.6	8.9	9.1	9.2	9.4	9.6	9.7
4.0	5.9	6.3	6.8	7.1	7.4	7.8	8.2	8.6	8.8	9.1	9.2	9.6	9.8	10	10	10	11	11
4.5	6.5	7.0	7.5	7.8	8.1	8.5	9.0	9.3	9.5	9.9	10	10	11	11	11	11	11	11
5.0	7.2	7.7	8.3	8.6	8.9	9.3	9.6	10	10	11	11	11	11	12	12	12	12	12
6.0	8.5	9.0	9.6	9.9	10	11	11	12	12	12	12	13	13	13	14	14	14	14
7.0	9.7	10	11	11	12	12	13	13	13	14	14	14	15	15	15	15	16	16
8.0	11	12	12	13	13	14	14	15	15	15	15	16	16	16	17	17	17	17
9.0	12	13	13	14	14	15	15	16	16	17	17	17	18	18	18	18	19	19

INSTRUCTIONS: Use this table for "probability" stock accounts, when deviations in number and size of demands must be considered.

1. Determine average number of demands during protected delivery time from ledger record.
2. Determine number of times per year stock is ordered from the Order Quantity Table.
3. At the intersection of the average number of demands and the number of times per year stock is ordered, find the Order Review Point in number of demands.
4. To express the Order Review Point in units multiply the number of demands by the average size of demands.
5. From 10 demands and up use standard tables.

FIG. 9 ORDER POINT TABLE FOR 1 STOCK OUT IN 5 YEARS

which in our experience has relatively much less weight than the number of demands, once derived, can be used in conjunction with the number of demands factor to obtain the total required protection.

The final problem in determining "when" to place restocking orders concerns delivery time. Although in Westinghouse we consider protection against deviations in delivery times in certain instances, we prefer to use forecasts. Therefore, we have established the practice that the Purchasing Departments shall keep the material control groups advised of current supplier delivery times. By the same token, the Manufacturing Divisions keep the Regional Sales Groups advised of current shipping times in the case of finished goods.

All of these factors can be integrated into a set of tables or programs for mechanical or electronic data processing. An order point table, which meets the particular needs of Westinghouse, is shown in Figure 9. Similar tables for various degrees of protection can be made using the set of curves in Figure 8. However, since these tables would not compensate for the additional variability introduced by variation in the size of demands, some compensations must be made. The simplest approach would be to develop tables having a theoretical high degree of protection. By observing the actual stock-out experience, one could approximately determine how much the theoretical protection should be reduced to obtain the desired degree of actual protection.

To calculate order points all the operator need do is enter the table with the number of times per year stock is ordered (based on the economical order quantity) and the number of demands during the expected delivery time (from past records), and at the intersection read the order point in number of demands. Next, the order point must be converted into units. This is the product of the average size of demands (use same past records as for demand) and the order point in demands.

* * *

Essentially this is the Westinghouse scheme for controlling stores inventories and customer service. With it management can get policy and quantitative decisions effectively executed, item by item, in a uniform manner. Further, when it is desired to make changes, these procedures provide a straight forward approach and make it possible to predict their effect in advance. With techniques such as these we can achieve "management control of inventories" and be assured that inventories are earning a maximum return on assets, the hallmark of a "growth company."

INTEGRATING OPERATIONS RESEARCH INTO A BUSINESS

Harlan D. Mills
General Electric Company

Introduction

The potential of scientific philosophy and practice as a participant in the operation of a business is probably little appreciated, even by our optimists. A limitation in the realization of this potential is the sophomore in all of us, both in regard to vision and complacency - it is probably safe to say that the misuse of ideas is fully as responsible for limiting human progress as the lack of them. In view of this, three preliminary notions are submitted as background for the remarks to follow.

The first notion is that a business ultimately exists solely for the purpose of individual human expression - the voluntary cooperation and organization of many individual resources and talents for a greater contribution as a whole and more satisfactory expression as individuals. While the operation of a business requires both ethical and intellectual principles, the appropriateness of the intellectual principles must stem directly from the ethical ones.

Secondly, there is an important distinction between the study of organized human efforts and the physical sciences - it is in the fact, in the first case, that the complexity of the subject matter is not independent of the methods of study. Whereas, say, the structure of an atom will "hold still" while more powerful methods for its study are developed, the very progress of science and technology demands an increase of specialization and complexity in the organization of human effort. It is easy to fall into the trap of appraising our abilities to study yesterday's problems with today's methods. Carried out to extreme, this mental lapse can produce an illusion that the intellectual activity of a business can ultimately be specified by formal science alone. We cannot afford to see only the wondrous progress of science and technology, with the more powerful methods of study it produces, while overlooking the advancing specialization and complexity in the organization of human effort this very progress demands.

Finally, the sober and continuous recognition of the inherent abilities of the human mind and intuition is submitted as a necessary means for the realization of the potential of science in business. This marvelous instrument is available which can compute, not only in black and white, but in all shades of gray. The human being must not be factored into business operations by science as an afterthought or a necessary evil; he is the most powerful logical instrument available and other scientific devices must be integrated around him.

The General Electric Approach

Our group in General Electric - the Operations Research & Synthesis Consulting Service - owes its existence to a charge from the management of the Company to explore the possibility of applying Operations Research in business. This stems from a desire to incorporate the philosophy and practice of Operations Research, if applicable, into the long-range planning and day-to-day work of the Company.

While this charge did not endow the individuals of the group with any special knowledge, it is affording the opportunity to study Operations Research, itself, from the point of view of its ultimate contribution to a business.

No amount of words or concept will themselves produce research - that is a property of people, their imagination and their industry. At the same time, the contribution of Operations Research work must finally hinge on the quality of the research itself. The remarks following presume and require a quality research effort. They are intended simply to set a framework in business and science in order to make effective contribution of research possible.

This framework is not in itself offered as a general answer for the problem of research in business operations. It reflects, first of all, our obligations to General Electric, the special characteristics of the Company, and its particular philosophy of managing. Needless to say, it is also limited by our own imagination, experience, and vision.

Some of the elements of the framework are believed to be innovations - the great bulk of them are well known, and have been ably expressed elsewhere. In general, the discussion will be concerned with what we believe to be the innovating ideas leading out of the common agreements of the field.

Research and Managing

There is common agreement that the potential of scientific method and the application of scientific results to business situations far outstrips the abilities of scientists and managers to cooperate, and hence, to contribute to business operations. The bottleneck lies in communication - on one hand, for the manager to tell the scientist what the situations are, and on the other, for the scientist to interpret the analyses of situations for the manager. This problem of cooperation has been an area of vital concern to our group.

No panacea has been discovered for this bottleneck. The only resolution proposed is embedded in an entire notion of research on business operations. It is not proposed as an easy answer, or even an answer at all. It is a way of thinking - not for scientists or managers separately but for scientists and managers together. This way of thinking represents an effort to structure Operations Research into the discipline of managing, as understood in General Electric, so that it becomes an integral part of the total work of managing the operation of a business.

Much of the structure has grown out of the consideration of the following key questions.

1. How do managers, themselves, develop abstract representations of situations and interpret representations into decisions, as well as how are representations logically manipulated?

2. What is the nature of situations which managers face, as well as what scientific methods are available to study situations?

3. What information is available to managers in situations as well as what information is necessary to specify optimal behavior in them?

It was not so much that answers were sought, but the focus the questions brought to the entire task. The answers, in general, are well known - the problem is what to do about them.

A basic premise evolved out of the first question: The use of scientific method is a present and necessary part of the intellectual activity of managing. More specifically, a pre-decision speculation on the part of a manager can be broken into the elements:

Objectives	-	preference orderings on future events
Assumptions	-	the appraisal of the situation itself
Strategies	-	various hypothetical courses of action
Expectations	-	hypothetical outcomes of the various Strategies, anticipated by means of the Assumptions, and evaluated in terms of the Objectives.

The contention is that in handling this set of elements - constructing and utilizing them - the manager uses scientific method. The limiting condition, aside from personal training and aptitude, is principally in the lack of two resources - time, and the luxury of procrastination. That is, the difference between research and this portion of the intellectual activity of managing is a matter of degree, not kind.

At the same time, these elements display the opportunity of research in business as that of augmenting and strengthening the logical connection between the inputs (Objectives, Assumptions, Strategies) of the representation and its output (Expectations) - that is, in reducing the Objectives, Assumptions, and Strategies to more primitive levels of belief and simplicity in refining the Expectations to sharper statements of consequences, and making the logical connections more precise and communicable.

The focus of the second question brought home the fact that managers face situations which, in terms of scientific standards, are tough ones indeed. The numerosity and intangibility of the factors they must consider is well known, as well as the requirements of making decisions with limited time for study. At the same time, these decisions must be made in an environment containing the whole range of uncertainties up to choices of other free wills.

Out of this came a pair of basic convictions. The first is that the bulk of the important managerial situations cannot be abstracted to a laboratory atmosphere. That is, as a rule, it is not possible to construct a representation of a situation which is both practically feasible for scientific manipulation and adequate, managerially, to stand as an independent description of the situation. The second conviction is that managing is distinctively a participation in a dynamic flow of events rather than a sequence of static situation resolutions - we speak of situations and their resolutions formally to pick out elements of the entire process for study and speculation, but this convention is an artifact only for thinking.

In regard to the third question, it almost seems that fate has played a monstrous trick on the managing profession - that the more important the situation, the less information there is about it. Controlled experimentation is out of the question in most situations, while many of the large business phenomena grind out their characteristics exceedingly slowly, if finely.

The magnitude and inherent complexities of business situations suggests the close association between managers and scientists, but more than mutual familiarity is sought in at least three contexts.

First, the research is regarded as managerial work - that is, the research team is considered a proper part of the entire management team, rather than a team of outside scientists. Its responsibilities are those of a responsible function of this overall team.

Secondly, the domain of activity is, in a real sense, the minds of operating managers, and not the operations of these managers. Neither the supervision of operations or the acquirement of vested interests in them on the part of the research team is intended.

Thirdly, the focus of activity is not in problem solving, but theory building - developing continuing stable frameworks for thought in the matter of managerial participation in a dynamic flow of events.

Briefly, some of the elements of the notion are:

- that research on business operations should be an integral part of the intellectual activity of the whole management team of a business,

- that the subject matter of the research is the creation of additional understanding, appreciation and usability of scientific method throughout this intellectual activity of the management team,

- that the object of the research is basic understanding of, and insight in, business operations rather than advice and recommendation,

- that the research be a continuing, permanent activity in recognition of a continuing, permanent activity of managing itself,

- that the business insist that the work actually be research, and not a super-operation or supervisory activity.

Some of the operational characteristics of such a research activity are visualized to be:

- separate and distinct managing activity with responsibilities as indicated above,

- research team (undetermined mix) of both managers (by experience) and scientists,

- organization only at business level,

- no complement of data collectors or clerks - the other managing functions to handle these tasks as in any normal course of business operations.

Some particular thoughts on this activity are:

- data collection, processing, and interpretation is a function of

the operating managers, not the research team. Understanding of concepts by operating managers behind this data utilization is the concern of the research team.

the utilization of high speed computation in management problems is likewise a function of operating managers, with a similar responsibility of developing understanding of concepts being the concern of the research team.

It is apparent that the real contribution expected from this activity is not in answers it gives, but in allowing greater utilization of the inherent potential of the entire management team to manage the business effectively. Obviously, our belief is that this is the best way to utilize research in business operations. However, we do recognize some risks inherent in the approach.

First, the General Electric manager is being asked to participate more actively in the business research effort than ever before. It is easier to give or receive advice than understanding. Secondly, the scientist is being asked to participate and understand the business at extremely close range with reality. The problem visualized here lies in the abstracting and imaginative powers of the scientist in such day-to-day work - that the scientist might become a victim of scientific provincialism and a slave of particular tools. This latter problem is part of keeping any applied research effort flexible and adaptable.

The first risk accents the attempt to structure a concept of Operations Research into the discipline of managing, and to display this structure to the managers of the company. The second points out the necessity for an adequate climate for research in a business, and indirectly right back to the essential understanding on the part of the management team as to what the research work is, what it can realistically do, and how it should be interpreted. Ultimately, both of these risks reside in individuals, managers or scientists, and in the soundness of the ideas offered to them.

Remarks on the Subject Matter of Operations Research

You will find no quarrel with us in the subject matter of Operations Research. That is, any contention will likely go by default as far as we are concerned. The feeling is that the opportunities to get at the large managing problems of whole businesses appear so great that we would welcome the chance to unload the borderline problems between Operations Research and Quality Control, Industrial Engineering, Communications, etc. Of course, we visualize Operations Research people working in these areas because of a void which requires filling, and similarly, people in other sciences will do Operations Research work.

A basic consideration in the subject matter stems from the conviction that the manager is distinctively a participant in a dynamic flow of events. It has to do with the approach to the research. We call it, for want of a better word, inferential rather than computational research.

In complex situations, a compromise must be made between the tractability and the realism of the representation. As the realism of the representation is increased there are often inferences which can be drawn, even though no formal resolution can be obtained. More and more, we visualize greater realism in representations and less complete manipula-

tion and resolution of them - cruder but more applicable results. We must not confuse crude results with crude thinking in this context. One might say the central limit theorem of statistics is a crude result, but it is certainly not the result of crude thinking. Rather, an impossibly difficult description of the exact state of affairs for samples of a given size from a given distribution would be the result of crude thinking.

Possibly a simple illustration would suffice for the point. In an actual situation facing a General Electric manager, a representation of the optimal course of action can be paraphrased as

$$\min \sum_{i=1}^{n} f_i(x_i) \qquad (1)$$

$$\sum_{i=1}^{n} x_i = 1 \qquad (2)$$

where the x_i are choices of the manager. By the method of LaGrange multipliers, it is easy to show that a necessary characteristic of the solution (x_i^o) is

$$f_i'(x_i^o) = p \qquad i = 1, 2, \ldots, n \qquad (3)$$

where p is an undetermined number. Physical considerations guarantee at least one solution to this set of equations - the only trouble is that n exceeds 500 and (if it mattered!) the f_i's are not linear; not even polynomials. Still, without a solution, the equations of (3) are a clue to the situation.

The boundary condition (2) states a condition which the manager creates in his day-to-day decisions. That is, any set of (x_i) chosen in normal operations satisfies (2). The approach was to consider in practice the actual x_i being used, and treat the $f_i'(x_i)$ in a manner analogous to a quality control chart. That is, let $z_i = f_i'(x_i)$, and plot z_i against i. If there is no variance, an optimum solution is at hand. In the continual process of making the decisions, repetitively, the manager is furnished guides by this pseudo control chart - he continually strives to get each point $f_i(x_i)$ to their current overall average. I.e., in any instance in a choice of an x_i, he knows whether to increase it or decrease it (and also has an estimate of how much). In this case, the manager, literally, is an integral part of the computing framework. The belief is that, more and more, close association will make this sort of process possible and effective.

Another foreseen opportunity of research is that of contributing completely outside the subject of the situation itself. That is, applying principles of organizing thought and speculation. Often, a complex situation is better understood by the simple process of structuring it in some way to provide perspective and a sense of completeness.

An example of this can be illustrated by a situation often arising. In analyzing a system of elements, it is easy to forget - or never bother to find out - just where the elements fit into the picture. In general, they can be divided into three classes:

- directly controllable
- directly desired
- neither of the first two.

Often, by tradition or business lore, some of the third category assume a false stature of one of the first two categories. If the system is a uniquely determined one, of course, all elements are controllable indirectly. Some elements of the third category may serve as guides, but they often end up as alleged panaceas. Sometimes this comes about because easily discernible concrete goals can be set for them, while the goal for a directly desired element cannot be so easily or concretely described. It may be known that this element should be minimized, say, but its minimum value is unknown. A case in point is the trio

Schedule	-	directly controllable
Cost	-	directly desired (minimum)
Turnover	-	neither of the first two.

Frequently, in practice, it is turnover which dictates schedule, rather than cost considerations. Many times, simply the recognition of this classification may contribute directly to the basic insight and intuition of the managers concerned.

Finally, in regard to the third key question, we visualize an opportunity of research to create stable ways of thinking in a business which will develop a purposeful history. It is not enough for the research team to look at a business as an integrated whole statically - it must look at the environment surrounding the business, and at that as a dynamic process. The sequences of acute situations must be described as manifestations of dynamic systems, so that a basic understanding of the participation of the business in its environment is developed.

For example, it is easy for a business to feel very proud of its profit position in a seller's market, and wonder ten years later why the industry capacity is over-extended to the detriment of its profit position at that time. That is, the boom and bust are part of the same situation over time and must be recognized and studied as such.

Regarding the subject matter, generally, it is easy to see that quantification is not a critical requirement with us. The boundary on one side is amenability to thought and communication - on the other, the ambitions and standards of the allied sciences. We would appreciate all the help possible - we need and solicit it.

It may well be that I have not been talking about Operations Research at all. This is ultimately a matter of taste. I have intended to talk about a science, which in cooperation with the presently established ones, can make the greatest possible contribution to human decision making in voluntary organizations. That is, a science which augments, rather than competes with what already exists. The problem of organized human effort is so great as it stands that competition for special areas seems foolish.

A CHECK INSPECTION AND QUALITY RATING PLAN

H. F. Dodge and M. N. Torrey
Bell Telephone Laboratories, Inc.

Another paper* has presented an over-all quality assurance plan used in the Bell System for assuring the quality of the various telephone products added to the telephone plant. This paper describes in some detail one part of this over-all plan, namely, the check inspection and quality rating plan that is used for products of discrete articles (often referred to as "apparatus") manufactured within the Bell System in relatively large quantities. Some modifications are necessary for the more complex products of wired equipment units that are produced in relatively small numbers.

Under the plan, inspection is performed on the manufacturer's premises by an inspection agency of the customer. For each class of product a series of small samples are selected during the course of the month from the flow of finished product en route from the manufacturing department to merchandise stock - and the customer. The amount of inspection is small compared with that of usual lot-by-lot inspection plans.

Individual units in each sample are inspected for engineering requirements and workmanship items and any defects found are noted. Defects are classified according to their seriousness as Class A, B, C, or D and each individual kind of defect separately is subject to a nonconformance criterion based on its seriousness and the sample size. If the number of defects for any characteristic exceeds the criterion, a second sample at least twice as large as the first is inspected. If the criterion is still exceeded, the lot represented by the sample is designated nonconforming. Except under closely defined conditions, nonconforming lots are returned to the shop for correction. Thus the nonconformance procedure imposes a limitation on quality with respect to individual characteristics.

To complete the general plan a second control is provided with respect to over-all quality. Defects are assigned demerits according to their seriousness and the over-all quality for a given month, as evidenced by the cumulative inspection results for that month, is given a demerit rating. This is plotted on a monthly control chart which shows quality relative to the standard level that has been established for that product. Significant departures from standard quality serve as a basis for corrective action on the process.

- -

*E. G. D. Paterson, "An Over-all Quality Assurance Plan"

AN OVER-ALL QUALITY ASSURANCE PLAN

E. G. D. Paterson
Bell Telephone Laboratories, Inc.

Quality Assurance embraces much more than inspection or quality control, or, as it appears to be used in much of the literature, a combination of these. Quality Assurance is, rather, a procedure by which the customer-user is continually assured of product whose quality is at a level which he reasonably expects. Thus it is primarily a function in behalf of the customer, and the quality levels embodied in the Quality Assurance plan may differ from those specified in the bare design requirements covering the product. At least, this is the significance of the term Quality Assurance as applied in the Bell System for the past thirty years.

Defined in this manner, Quality Assurance to be effective must

(1) Establish quality standards which, from the point of view of the customer, represent product which is satisfactory, adequate, dependable, and economic.

(2) Continuously evaluate product quality in terms of these standards.

(3) Initiate measures to prevent any significant departures of product quality from these standards, and

(4) Provide continuing evidence as to actual product quality.

Successful fulfillment of these four functions requires the exercise of separate but closely interacting activities under the following headings:

1. Setting Quality Standards
2. Preparing Inspection Procedures
3. Disposing of so-called Non-Conforming Material
4. Handling Engineering Complaints
5. Conducting Quality Surveys
6. Analyzing, Evaluating, and Reporting Quality Results

The paper discusses these six activities, their interrelation, and the type of organization and personnel required for their prosecution.

SELECTIVE ASSEMBLY

R. B. Murphy
Bell Telephone Laboratories, Inc.

In assembling two piece-parts with additive characteristics a major concern is the statistical behavior of the resultant characteristic in the assembly. This behavior is partly determined by the probability distribution of the piece-part characteristics and partly by the statistical nature of the method of assembly.

Various methods of assembly, in particular of selective assembly, are compared in terms of the variance of the assembly characteristic and also the "loss" in scrapped piece-parts or defective assemblies. Most of the results concern normally or uniformly distributed piece-part characteristics, while some pertain to less special cases.

METHODS IMPROVEMENT THROUGH QUALITY TECHNIQUES

Blair E. Olmstead
The Prudential Insurance Company of America

Quality Control applications in the Prudential Insurance Company of America are handled by the Quality Improvement Staff. Our group is one of many in the Planning and Development Department. The Department is responsible for over-all plans for future operations and plans for reduction of cost and improvement of service in both Field and Home Office. It contains groups operating in special areas of methods, personnel, economic and market research. Our activities, however, are not restricted by this specialization. Operations are often examined by members of two or more groups. This can, in some respects, be likened to the Operations Research approach.

The stock-in-trade of the Quality Improvement Staff is the Quality Improvement Program. We call it Quality Improvement rather than Quality Control since it is aimed at improving the work of clerks rather than controlling them.

At the invitation of line management, our group will survey an operation, design a suitable system of sampling the work and install a review of the quality of this sample. The review, however, is not conducted by a member of the Quality Improvement Staff, but rather by a clerk who would normally perform the operation under consideration. In fact, we go one step further and rotate all of the clerks in the quality reviewer position.

The Quality Improvement Staff combines, computes, publishes and analyzes the results of the quality review. The analysis, along with our recommendations, is given to the line management and supervision.

Through the analysis of the results of a Quality Improvement Program's quality review, management receives facts about the process that were unavailable to him before. In addition to the over-all accuracy, he is able to have at his fingertips data about the relationship of one type of error to another. The Quality Improvement Staff often collects information regarding the cost of errors made during the operation and evaluates the inspection system in the light of initial accuracy, inspection efficiency, and error cost. The Quality Improvement Program also has a predictive aspect. It supplies the management with information regarding anticipated accuracy under various conditions.

A Quality Improvement Program, while designed and installed by a staff group, is basically a tool of the local line management. It is their program, and the Quality Improvement Staff is, essentially, supplying a management service.

Our position, part of the Planning and Development Department yet working closely with line operations, gives a two-directional flow to our activities. When, during the course of our quality control applications, we encounter a situation which we feel could stand examination by a specialist from some other area, we call him in. Conversely, other specialists and line management make use of us whenever a problem involving sampling, clerical accuracy, etc. arises from their work. The following examples of this two-way flow have been selected to show the interplay of forces in obtaining improvement in clerical methods, and the part of

quality techniques in each.

Digit Grouping

Almost every item handled in the insurance business has a number, usually a policy number, associated with it. This constant presence of an identifying number has been of considerable assistance to us in sampling. These numbers are usually six to nine digits in length, or longer than the eye can comfortably grasp in one glance.

In one of our Quality Improvement Programs, which covered a typing operation, significant differences were observed in the accuracy of the copying of eight digit policy numbers from various source records. The Quality Improvement Staff man handling this particular Quality Improvement Program, analyzed these source records and found that those with unseparated digits had more errors. In order to find if this were peculiar to this particular group or would have application throughout the Company, he also consulted the Testing Unit who developed a test of the accuracy with which grouped and ungrouped multi-digit numbers could be grasped. These tests showed conclusively that the separation of long numbers into groups of two and three by blank spaces, hyphens or vertical lines had a marked effect on improved accuracy and production. After a standardized practice of digit grouping had been determined, the Forms Control Group set in motion the wheels to revise forms throughout the Company to provide for digit grouping.

In this example, we have the results of a Quality Improvement Program on a relatively minor operation inspiring a large-scale examination of a practice throughout the Company. The examination of this practice (testing of the various digit presentations) was conducted by the appropriate group, the Personnel Research Testers, and the corrective machinery, forms revision, carried through by the appropriate Planning and Development Agency.

Records Conversion

In this illustration, line management, having decided that the time had come to examine the condition of some of their source material, called on the Quality Improvement Group first to supply them with facts in the form of an estimate of the total number of entries to be converted, and second to set up and maintain the inspection procedure with the aim of obtaining an acceptable product with a minimum of cost.

Some of our older records were kept in a form which has become unwieldy. In this particular instance, large ($7\frac{1}{2}$ pounds) books containing policy particulars for almost 3,000 policies each, reached the point where only 25 per cent and less of the entries were for policies currently in force. It was, therefore, decided that the time had come to convert the remaining "live" entries in these books to cards.

The Quality Improvement Group was asked to set up the inspection procedure, so that the desired degree of accuracy in converting these records could be attained with a minimum inspection cost. The conversion was made by typing the information from the books to cards. These cards and books were then compared until the desired accuracy level was reached, as indicated by the acceptance sampling plan.

Because of the unique (once completed, it was completed for all time)

nature of the job, the typists were obtained from various points throughout the Company as they were available and there was considerable turnover among the typing group. We soon found that we had a tri-modal accuracy curve for the typists. A small number were found to be extremely accurate. The bulk of the typists produced work which required one or two inspections to clean out the errors and a small, but distinct group of typists required at least two, often more, inspections before their work could be deemed acceptable. We were able to incorporate this tri-modal curve in our inspection procedure by comparing the cost of our sample review with complete inspection and the probabilities of acceptance associated with each of the three modal groups.

As the conversion progressed toward the older books and the proportion of "live" entries became smaller, one error, probably the most serious, was found more frequently in the typing. This error was omission of an entire card. When this happened, the inspection cost increased, since the inspector was asked to create a card whenever it was omitted. To reduce the frequency of this error, we added a processing step before the typing operation. This step consisted of leafing through the books and inserting blank cards wherever there was a "live" entry. This card insertion step gave us a 15 per cent increase in typist production, but cost us a total of 25 per cent in total first work (inserting and typing) staff. However, it was more than justified because of a 20 per cent decrease in the total (including quality review) inspection time.

<u>Central Recording</u>

The Prudential recently changed its dictating facilities from individual desk machines to a central recording system which involves several telephone-like instruments located on the dictators' desks which connect to banks of recorders at the transcription center.

Our group was called upon for statistical services in a sample study of the traffic pattern of recording, equating the various peak periods to reduce fluctuations at the recording and transcribing end, and determining the optimum dictating stations to recorder ratio. This ratio involved such factors as the anticipated "collision rate" during normal and peak operation, the cost under the several different hookups that were available and means of detecting overloading before it had become serious.

In this instance, the specialists in Office Equipment had the prime responsibility for determining whether central recording was desirable and called on the Quality Improvement Group for assistance in obtaining the facts for their decision.

<u>Check Typing</u>

One of our largest volume check-typing operations, involving the time of 12 people, was being considered for a change in equipment. In place of our practice of first typing checks and then preparing accounting records, it was proposed to use a new machine that would prepare a paper tape at the same time the check was typed. From this tape, the accounting record, in the form of a punched card, would be run off.

The accuracy of the finished product, checks after comparison, was quite high and deemed satisfactory. The question to be resolved, however, was that of the initial accuracy with which checks were typed, since errors picked out by the comparer, or by the typist herself after a

certain point in the procedure, would no longer be correctable by merely voiding the check and writing a new one. It would now be necessary to void all of the accounting records to be prepared by the paper tape.

Our first objective was to measure the current accuracy with particular emphasis on the types of errors which would cause greater difficulty under the proposed system of check typing and record preparation. After we had determined this accuracy, we found that one kind of check was less accurately typed than others. However, it appeared that the new machine would not encounter undue difficulty in operation because of typing inaccuracies. Nevertheless, a Quality Improvement Program was installed for the entire operation and in a due course of time (in this instance, only a few months) the quality of the typing, particularly the typing of the class of check having the lower initial accuracy, had stabilized at a highly satisfactory level.

The scope of the Quality Improvement Program was then curtailed. A review of a very small fraction of the work, only sufficient to detect relatively large month to month variation, was substituted for the close analysis made during the earlier stages of the program.

SUMMARY

Quality Improvement, as Quality Control is known in the Prudential, is being applied over a wide range of clerical activities. It functions successfully, both as a tool of line management and in conjunction with other staff groups, to improve the methods of clerical operation.

TEST STANDARDS

J. S. Wiberley
Socony-Vacuum Laboratories
Brooklyn, New York

I should like to take the next few minutes to explain how one industry - the petroleum industry - sets up and maintains test standards. The uniformity of petroleum products is controlled largely by means of physical properties, such as boiling range or viscosity. These properties are measured by tests. Because petroleum is a very complex mixture and because buyers and sellers cannot wait, the industry has adopted a number of standardized methods - some of them rather arbitrary in their details - which are recognized for testing petroleum products. Without these test standards, sale of the products would be chaotic. For instance, these tests define gasoline and distinguish it from kerosine. Tests are, of course, only a part of the picture. Ultimately, performance determines what is a good product. Tests are none the less an essential tool in the maintenance of quality.

The organization by which these tests are standardized is Committee D-2 on Petroleum Products and Lubricants of the American Society for Testing Materials. The first test was made a Standard over thirty-five years ago. In 1954 Committee D-2, working through its technical committees and research divisions, published one hundred sixty tests. Each year a manual is published having all of these methods in up-to-date form. This manual is the handbook for all large-scale buyers and sellers of petroleum products. Producers and consumers share in the selection and standardizing of the methods. All of the work is volunteered, yet, so essential is this function, willing hands are always found.

Standardization is a deliberate and slow process. It may take five years for a test to be so recognized. First, the test must be proposed. If it is a brand new test, the technical committee responsible for the particular product must agree that there is a commercial need for such a test and recommend that the related committee section investigate. These sections are small, usually consisting of ten to twenty members. The members try the test in their own laboratories on some cooperative samples and analyze the data statistically. If the test is satisfactory, it may be submitted to the parent committee and D-2 for publication as a Tentative. Usually, however, some revisions are necessary before satisfactory results can be obtained in many laboratories. In the meantime, the test may be published for information in the back of the D-2 Manual. Opinions are solicited.

After it has been thoroughly tested and meets with unanimous approval through the hierarchy of committees, it appears as a Tentative with a number designation. After one or two years as a Tentative, it may become a Standard if

there are no objections and no anticipated changes in the procedure. After the method is widely accepted, D-2 may recommend and the American Standards Association (ASA) may approve its adoption as an American Standard.

This is the manner by which the producers and consumers of petroleum agree on tests which will apply to specifications. These tests cover most commercial needs. However, there are interim and local needs where other tests must be standardized. For example, each petroleum company needs methods for control and research, methods which will be recognized and be run in the same way in laboratories within that company. Most large companies have a book of such methods. Socony-Vacuum has a loose-leaf methods book to which revisions and new methods are added several times each year. These methods may be proposed by anyone in the company. The draft is reviewed by two four-man committees, one at the Research and Development Laboratory and one at the Technical Service Laboratory. These committees see to it that each method is clearly and completely written and that it is technically sound. In some important cases, cooperative work is undertaken by two or more laboratories to check the precision of the method. When it has been revised and approved by the committees and by laboratory management, it is sent out to holders of the methods book.

Throughout any standardizing work, precision is something that comes up constantly. If a method is to be standard, it must yield consistent results. An operator should be able to repeat the test on the same sample and obtain the same, or nearly the same, result. (In the petroleum industry this kind of precision is termed repeatability). Different laboratories, e.g., the seller and the buyer, should likewise be able to get similar results. (In the petroleum industry this is called reproducibility). The need for good precision is self-evident. Standard methods generally have some statement as to how closely duplicate results should check. Nevertheless, there is still a great deal of confusion today because of the loose way that precision is often stated. A test is said to be "good to ten percent" without any amplifying remarks as to how often, or under what conditions, it will be "good" to this amount. A.S.T.M. has a committee (E-11) working on precision definitions and problems, and D-2 within A.S.T.M. has its own committee on precision of methods. This latter committee, under the chairmanship of Fred Tuemmler of the Shell Development Company, has recently published a recommended practice for applying precision data (2). This practice arbitrarily defines the way in which precision should be reported. Gradually, this practice is being recognized and the various subcommittees of D-2 are attempting to revise their precision statements to conform. No doubt there are other ways to state precision that would be more attractive to some; in fact, D-2 has been criticized for not labeling their definitions "D-2 repeatability" or "D-2 reproducibility" to emphasize that these terms are being used in a restricted

sense (3). The restrictions briefly are these. It is assumed that in practice two results are compared. If the results were obtained by a single operator in one laboratory (generally with the same apparatus and within a small interval of time) these two results should agree within the repeatability 95% of the time. If they were obtained by two operators in two different laboratories, they should agree within the reproducibility 95% of the time. Fortunately, there has been ready acceptance of the arbitrary 95% confidence. The Institute of Petroleum, the British standardizing body, set the pattern that was followed (4).

Committees and particularly inventors of methods have tended to think of usual deviations, with confidences of only 50 or 60%. So the new definition has raised the figures. Also, deviations have been measured from a mean rather than between two random results. This, too, makes the figure bigger - sometimes frighteningly so. I have seen several cases where the precisions were considered acceptable until they were translated into the new terms - then they were not good enough. Test operators are prone to ignore the probability curve and to find excuses for those occasional outlying results which should be expected. The D-2 practice at least has had the effect of wedging some statistical foundation under precision statements and is forcing many people to use statistical principles who never did before.

One deterent to the use of statistics is the time it takes to calculate standard deviations and probability limits in order to come up with an opinion that one might get just by inspection. For that reason, we are trying, in our company, to take some every day short-cuts where permissible. By way of illustration, let me describe a procedure we use to discover how well a given laboratory can check itself in normal operation.

In our laboratory all regular samples after testing go to sample storage for a set time. Some of these samples are selected at random and sent through for testing a second time. New containers and labels are used so that they cannot be identified with the previous samples. Twenty of these repeat tests are run. All results are then tabulated with first result, second result and difference (range) in three columns. The average range is obtained by adding the twenty differences and dividing by twenty. This average range is multiplied by 2.45 to obtain the repeatability of the test in the new D-2 terms. The arithmetic is very easy and close enough for the purpose. Precision figures are usually rounded off anyway. We have learned that one should not be too precise in quoting the precision of a test.

You are probably wondering where the 2.45 comes from. This is a combination of several factors.

$$\text{Repeatability} = \frac{\text{Range}}{1.128} \times 1.96 \times \sqrt{2}$$

The 1.128 is the factor for converting range of two into a standard deviation. The 1.96 allows for the possible deviation of one result from the mean 95% of the time. The square root of two is necessary because repeatability is measured between two such results; the two deviations are added but as a root mean square sum.

This same short-cut can be taken in reproducibility studies between laboratories. If, for instance, four laboratories engage in the work, then a range of four is taken (the highest minus the lowest result). Instead of 2.45 the combined factor would then be 1.35, since the factor for converting range of four is 2.059. Obviously, the range between highest and lowest will increase as the number of laboratories increases and the factor will decrease to compensate.

This discussion of precision may sound like a digression from the original topic of test standards. It was brought in in order to present some aspects which are new and still, to some extent, controversial. It is obvious that only a very few aspects of the broad subject "Test Standards" could be mentioned in such a short talk. It is hoped that some of these examples, taken from experience in the petroleum industry, will apply in other industries. I have tried to explain -

1) How the petroleum industry writes test standards.
2) How a petroleum company writes test standards.
3) How the petroleum industry defines precision of tests.
4) How laboratories can check their precision and quickly estimate the result.

<u>References</u>

1. ASTM Standards on Petroleum Products and Lubricants, ASTM Committee D-2, Philadelphia, 1954.

2. Ibid pg 896.

3. Bicking, C. A., ASTM Bulletin, September, 1954, pg 48.

4. Standard Methods for Testing Petroleum and Its Products. Institute of Petroleum, London, 1953, pg 715.

APPLICATIONS OF REGRESSION ANALYSIS
TO STEEL PLANT PROBLEMS

Donald S. Leckie
Republic Steel Corporation

Beginners in the use of Statistical Methods are frequently puzzled by the word "regression." After reading up on the term and working a little bit with the associated methods, they begin to realize that the word is in some manner concerned with a specific process of estimation. This specific process deals with the estimation of the numerical value of one variable through the known value of another and associated variable. This estimated value is only an average or expected value. The estimated value will seldom occur precisely, but some probability statement can usually be made as to its accuracy. Such a statement implies that a value will fall with an accompanying probability within a specified range about the expected value. These numerical estimates can be made from lines or curves or their equivalents expressed as equations. Such lines and curves have come to be called regression lines or regression curves. Others refer to this process by different names, such as least squares or correlation.

Regression analysis can be concerned with the attempt to estimate the value of one variable from the known value of some other variable, in which case it is called simple regression analysis. It may also be used to develop an estimate from a combination of the known values of different variables, in which case it is called multiple regression analysis. Regression analysis can be used to establish either linear or curvilinear relationships.

In this paper the regression analyses discussed will be linear, but examples of both simple and multiple regression will be given. This is done because experience has indicated that the first attempts at analysis of steel plant data should be made with the assumption of linearity. If it be found that this assumption is not valid for the problem, then the more complicated methods of curvilinear analysis may be tried. These latter methods could easily be the subject of a paper in themselves and will be disregarded here.

The paper will not dwell on formulas and methods of computation. Many recently published textbooks and articles may be consulted if the reader is interested in further development. Rather, the concentration will be upon two problems, each of which has some unusual facets which are not generally exhibited in textbook models.

AN EXAMPLE OF SIMPLE REGRESSION

STATEMENT OF THE PROBLEM

The first discussion will be focussed upon a simple regression analysis of data. When the discussion is finished you may agree that the word "simple" is not applicable in its non-statistical meaning. This problem arose in connection with the attempt to estimate the spread in sulphur analysis of billets rolled from a given heat of resulphurized steel, from the ladle sulphur analysis of that heat.

For the benefit of those not familiar with the above terminology, a resulphurized steel is one to which sulphur has been added at the end of the melting process to bring the final sulphur content of the finished melt up to a higher than normal level. This resulphurization improves the machinability of the products produced from the steel. Such steel is considered to meet or not meet specification on the basis of the "ladle" analysis, which is performed on samples taken during the pouring of the molten steel into the ingot moulds.

Unfortunately, due to the nature of the solidification process, the sulphur does not remain homogeneously distributed throughout the solidified ingot. Some segregation of sulphur results, and if a sample is taken from a billet rolled from such ingots, the sulphur content in the sample will vary somewhat from the ladle sulphur analysis. Such a sample might be analyzed to determine the association of a certain billet with a specific heat. To state the test statistically, this analysis is made for the purpose of testing the hypothesis that the steel which yielded such a sample is in fact from a heat of such a ladle analysis or grade.

Two problems can be formulated from the same set of data which is comprised of ladle analyses, together with the associated billet analyses of the same heats.

Problem 1. For a specific ladle sulphur analysis from a given heat, what average and what limits may be expected for billet analyses on that heat?

Problem 2. For a given sulphur billet analysis, what average and what range of ladle sulphurs could be expected to yield such a billet analysis?

It is important to note that the two problems are not necessarily converse to each other. In general, a regression solution for Problem 1 will not be a solution to Problem 2, and vice versa. This results from the solution of the problem. The solution provides a line which either minimizes the sum of squares of deviations of one variable, or of the second, about the regression line. Thus, in general, a different equation is required to answer each problem, although once one equation is known in the linear case to which this discussion is restricted, the other may be computed from it.

TECHNOLOGICAL ASPECTS

Those familiar with resulphurized grades of steel already had a general knowledge of the behavior of such data. Experience had indicated first of all that a greater spread in billet sulphur analysis could be expected when the ladle sulphur was high; a smaller spread could be expected to result from lower ladle sulphur heats. Secondly, it was the concensus that the deviations from ladle analysis of billet analyses higher than the ladle analysis were greater than deviations of billet analyses lower than the ladle analysis. There were sufficient theoretical arguments to support these contentions. These forecasts were warning signs to the statistician of traps to be avoided. If it were true that the range of billet analyses increased as the sulphur content of the heats increased, then the data would not have uniform variation about the regression line for the range of ladle sulphurs

encountered. If the second prediction were true, it would be entirely possible that the data would be "skewed" with respect to the regression line rather than normally distributed about it.

The usual textbook explains the procedure to follow if the data are normally distributed with constant variance about a regression line. Under such circumstances, a number called a "standard error of estimate" can be computed from the data, and this number used as a standard deviation of the data about the regression line. If the data were, in fact, normally distributed with constant variance about the regression line, it would then be expected that a zone having a width of three standard errors of estimate above and below the regression line should contain 99.73% of the data.

If either the condition of normality, or the condition of constant variance were not met, then the preceding statement need not be true. In this case the statisticians were singularly unblessed, because the data proved to be neither normally distributed nor to have constant variance. The data analysts were, therefore, forced to improvise a set of limits which would describe the actual behavior of the data. Although there may be more precise ways of greater mathematical elegance for doing this job, this illustration is advanced as a practical approach to the problem, and as an illustration of a fundamental viewpoint of regression.

SOLUTION OF THE PROBLEM

Fig. 1 illustrates the scatter diagram of the subject data. The values of the ladle sulphur analyses are plotted along the abscissa and the values of the billet sulphur analyses are plotted along the ordinate of the graph. The linear regression line, or computed line of best fit, is shown. It is possible to compute a standard error of estimate, whether or not its value is meaningful to the problem. Limit lines have been drawn on the chart at a distance of three standard errors above and below the regression line, and parallel to it. It is evident in the figure that the lower limit is too loose, while the upper limit is too tight and excludes a considerable portion of data. This results from skewness of the data with respect to the regression line and is a verification of predictions. Close examination of the figure indicates a tendency for the scatter of the data along the regression line to increase with increasing ladle sulphur. This is more easily seen in Table 1.

The data plotted in Fig. 1 represent 2595 associated analysis values which were available. This provided plenty of data for subdivision into groups with respect to small increments of values of ladle sulphur analyses. Thus, the billet sulphur analyses for ladle analyses ranging from .080% through .084% sulphur were collected in the first subgroup. The second subgroup includes those billet sulphur values from heats ranging from .085% through .089% ladle sulphur, etc.

The frequency histograms for the ten subgroups of billet sulphur analyses were constructed and these verified the fact that the variance tended to increase with increasing ladle sulphur, and also indicated the skewness of the distributions. Statistics computed for the ten frequency distributions included the average, standard deviation, a_3 and a_4. These values are tabulated for the ten frequency distributions in Table 1. By way of definition, a_3 is the third moment about the mean

FIGURE 1

BILLET VS LADLE SULPHURS

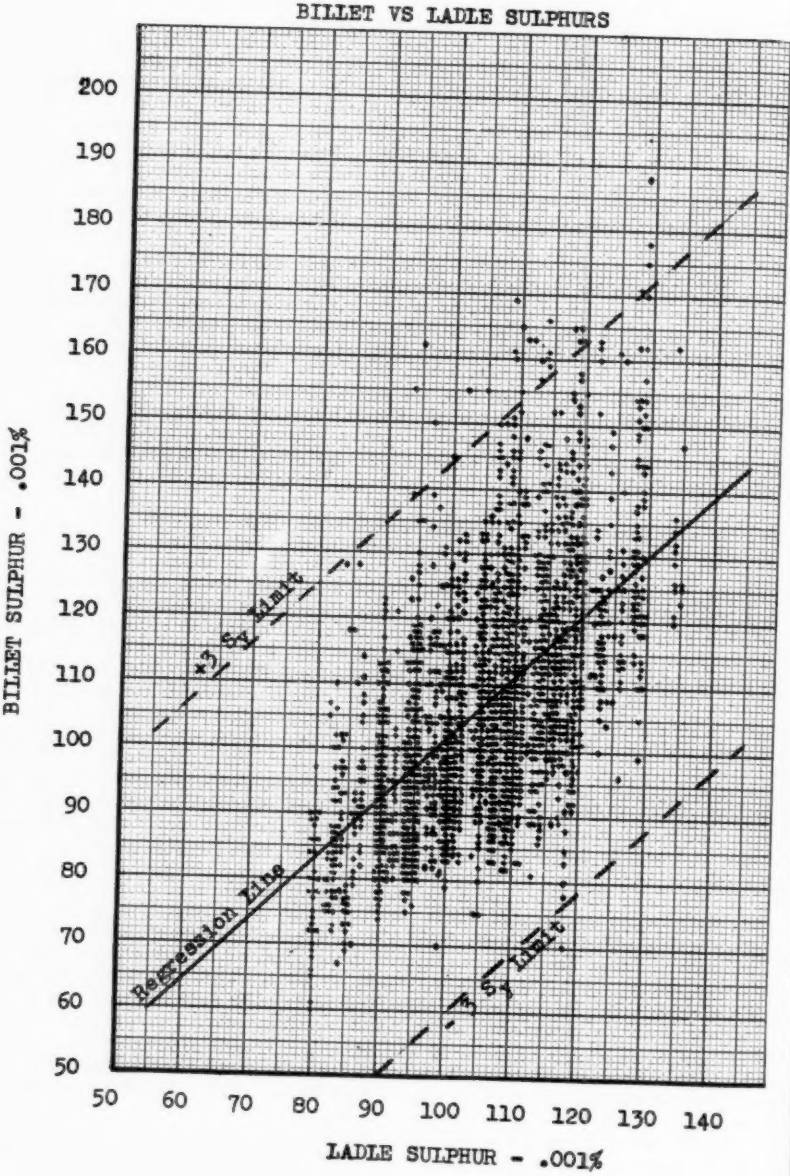


TABLE I

STATISTICS COMPUTED FOR FREQUENCY DISTRIBUTIONS OF BILLET SULPHUR ANALYSES FOR RANGES OF LADLE SULPHURS

CHECK SULPHUR FREQUENCY DISTRIBUTION DATA

Ladle Analysis Sulphur Range	Number of Data	Average	Standard Deviation	$V = \frac{\sigma}{\bar{X}}$	a_3	a_4
.080 - .084	108	.0855	.0111	13.0%	+.6533	4.03
.085 - .089	67	.0891	.0124	13.9%	+.6625	3.34
.090 - .094	294	.0942	.0115	12.2%	+1.2286	7.11
.095 - .099	305	.0996	.0135	13.6%	+1.0800	4.69
.100 - .104	294	.1024	.0130	12.7%	+ .7322	3.57
.105 - .109	596	.1089	.0148	13.6%	+ .8319	3.75
.110 - .114	356	.1146	.0152	13.3%	+ .7318	3.68
.115 - .119	328	.1168	.0158	13.5%	+ .5472	3.42
.120 - .124	134	.1234	.0136	11.0%	+ .6931	3.33
.125 - .129	99	.1309	.0180	13.8%	+ .9021	3.48

NOTE: a_3 = (Third Moment About Mean) ÷ (Cube of Standard Deviation)

a_4 = (Fourth Moment About Mean) ÷ (Fourth Power of Standard Deviation)

For Normal Distribution, $a_3 = 0$, $a_4 = 3$.

divided by the cube of the standard deviation, and is a measure of the skewness of the distribution. A positive value of a_3 indicates that the frequency distribution has a long tail extending in the direction of higher billet analyses. The a_4 value is the fourth moment of the distribution about the mean, divided by the fourth power of the standard deviation. It is a relative measure of peakedness or flatness of distribution, especially when the distribution is not skewed.

The a_3 and a_4 computed values provided further evidence that the normal frequency distribution was not a satisfactory approximation to the data in this problem. There does exist a system of equations called the Pearson System, which permits the fitting of frequency curves of various shapes. The normal frequency curve is a special type in the Pearson System. Unfortunately, tables of areas under various type Pearsonian curves are not readily available and, in many cases, do not exist. (The normal curve is an exception to that statement). Tables are available, however, for the Pearson Type III curves in limited ranges. (Ref. 1.) Fortunately, the data in Table 1 for several of the subgroups indicated that the Pearson Type III curves would be satisfactory approximations. Other subgroups indicated that while the Type III curves would not be quite as close approximations to the data, at least they would account for some of the skewness exhibited.

The tables for the Type III curves were utilized to fit curves to the data in the conventional manner. Such curves proved to be satisfactory from the standpoint of approximately describing the behavior of the data within subgroups. Considering the ten subgroups as samples from a Pearson Type III population, one could not expect the statistics for the ten subgroups to have values exactly equalling the parameters for the assumed Type III curve of the population.

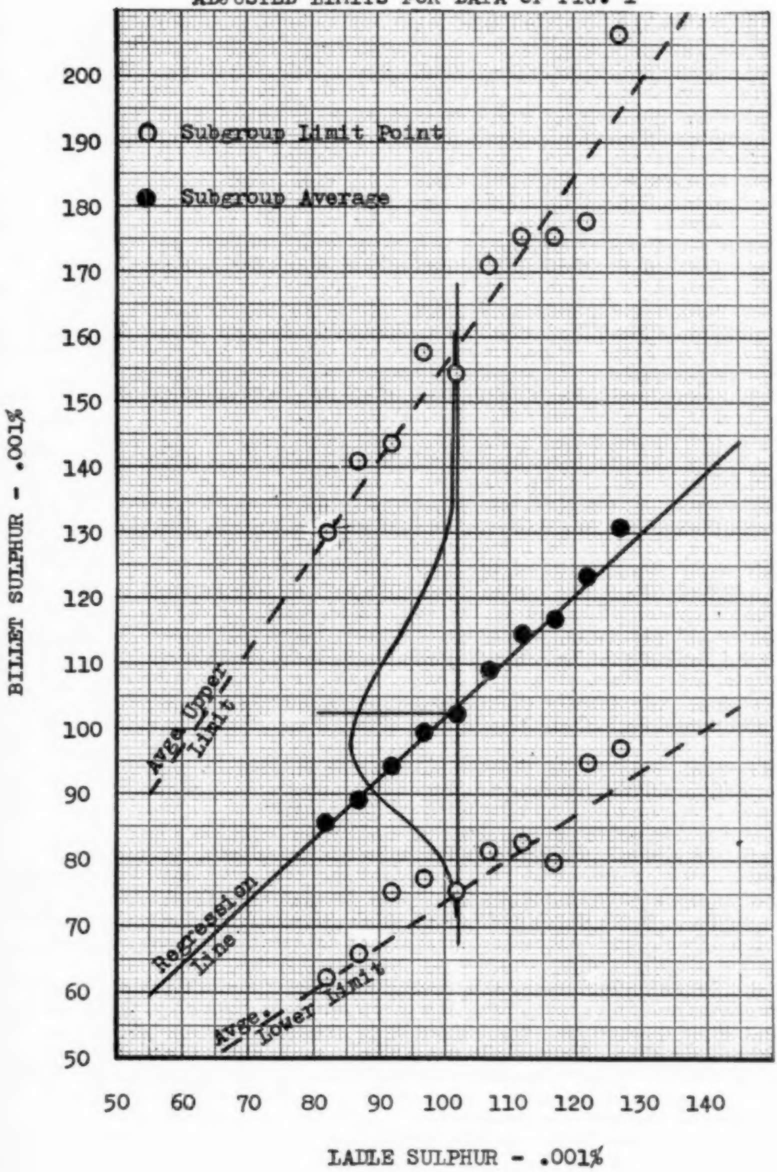
The next question for consideration was the one of limits for individual points about the regression line. If the data were distributed normally, with constant variance about the regression line, this would be a simple matter. Due to the nature of our data we had to consider other solutions. One method would be to find points for each of the ten skewed distributions such that 0.13% of the area is excluded under each tail of each of the curves. Such limits would compare, in principle at least, with the 3-sigma limits commonly used in quality control work.

Fig. 2 is an illustration of the conclusion of the problem. The regression line is shown, together with the average billet analysis for each of the subgroups. The upper and lower limit points for each subgroup are also plotted. As described above, these points are such that 0.13% under the tail of the distribution fitted to the data in the subgroup will be excluded by the points. The dotted lines are limit lines which were computed to fit the subgroup limit points. A typical Type III distribution for one of the subgroups has been superimposed on the figure.

It is interesting to check the expectation that 0.26% of the individual points will be excluded by the limits computed in this manner. This would mean that for 2595 points, approximately seven points could be expected to be excluded. Actually, three points exceeded the upper limits and nine points were below the lower limit, for a total of 12, or 0.46%. From a practical standpoint, this is considered entirely satisfactory. The fact that a few more points actually exceeded the limits than were predicted can be explained from many standpoints, chief of

FIGURE 2

ADJUSTED LIMITS FOR DATA OF FIG. 1



which would appear to be the assumption that the distribution of the data could be represented by an average type of Pearson Type III curve.

It must be admitted that there may be more precise and elegant ways of working this problem, but it is believed that the method presented did yield results which met the purpose of the investigation. This method had the advantage of approaching the concept of regression from a definitive, almost intuitive standpoint, rather than from a formal, formula-substitution type of solution. The results did confirm the metallurgists' general concept of the behavior of the data, and did serve to quantify their concepts.

Again, it should be stressed that this problem was to describe the behavior of billet analyses for a given ladle sulphur analysis. If it were desired to predict the range of ladle sulphur analyses which might be expected to produce certain billet sulphur analyses, an entirely new computation would have to be made. The new regression line could be computed from statistics already developed, but new frequency distributions of ladle analyses for arbitrary ranges of billet analyses would have to be made. The results presented above are not valid for this new purpose and this principle is one of the important principles of regression to be remembered.

The foregoing has been an example of simple regression analysis. To repeat, the word "simple" proved to be something of a misnomer in consideration of the problem's complexity. It is an unfortunate fact that several factors may influence the value of a related variable. In the previous example, the investigation might be extended to determine the effect of the portion of the ingot represented by the billet analysis, or the effect of the position of the ingot in the sequence poured. The consideration of the three factors, ladle sulphur analysis, sequence number of the ingot, and location in the ingot represented by the check, would necessitate an analysis such as <u>multiple</u> regression to estimate their independent effects, if any. Other methods, such as analysis of variance, or analysis of covariance could also be used under the proper circumstances.

AN EXAMPLE OF MULTIPLE REGRESSION

STATEMENT OF THE PROBLEM

The second problem is concerned with the attempt to measure the difference in effect of limestone from two different quarries upon the performance of open hearth furnaces. Some of the restrictions on the experiment reflect the difficulty which is encountered when an attempt is made to apply statistical principles to steel plant methodology.

In the first place, the quantity of limestone available from quarry X restricted the number of heats which could be charged with it. Rather unlimited quantities of limestone from quarry Y were available, but unfortunately the point of the experiment was to measure the effect of limestone X, relatively unknown, with limestone Y, which was the standard limestone charged. In consideration of the variation generally encountered in most open hearth data, there existed the distinct possibility that the amount of experimentation permitted by the material restriction would not admit clear-cut decisions.

It must be recognized that the data collected for an individual heat of steel is representative of a large quantity of material (200-300 tons of steel) worth a considerable sum of money. Such data has much greater value associated with it as compared to a measurement on a bolt, for example. Even if the variations are large and possibility of inaccuracy exists, such data are deserving of exhaustive statistical tests before any conclusions are made. It usually requires about twelve hours to produce one heat of steel. The collection of a large amount of data would, therefore, require a considerably extended time period. However, those familiar with the process will probably recognize that it is not desirable to continue such experiments over too long a period of time, due to increasing disinterest of employees, long-term trends in shop conditions necessitating changes in furnace practice, etc.

Thus, an experiment had to be devised to balance the above two restrictions, among others. In addition, due to conditions with which every operator is familiar, it is frequently not possible to adhere strictly to a schedule which has been designed for the benefit of the investigator. This is another source of complication. Also, it is not likely that any two heats will be produced under exactly the same, or even similar conditions. This will lead some to wonder if perhaps such an experiment should be tried at all, and if small scale laboratory experiments might not be more desirable. Actually, regardless of the extent of small scale laboratory and pilot plant experimentation, the propositions still must be evaluated with operating equipment under operating conditions before any conclusions can be drawn with some degree of certainty. Thus, an experimental plan was necessary which compromised with some of the above limitations.

It was finally decided to make as many heats as possible from the available supply of limestone from quarry X, and to make an equal number of heats using limestone from quarry Y. The limitation of available material prevented the length of time for the experiments from being such that long term trends would have an important effect upon the results. Also, it was planned to alternately make a series of four heats with limestone X and then a series of four with type Y limestone. In this way shorter trends and cycles would have the opportunity to influence both groups of heats to approximately the same extent. Making alternate heats with the different limestones would have accomplished this with even greater efficiency, but it was expected that a slightly different slag system would result from the use of limestone X, which effect would tend to be concealed by the slag remaining on the hearth of the furnace after the heat was tapped. It was, therefore, desirable to make a short series of heats from alternate types of limestone, since in this way a difference in slag might be more easily detected and the effect on the furnace bottom evaluated.

Since open hearth furnaces are quite individualistic in their performances, it was decided to limit the test heats to a single furnace in the shop. This removed a source of variation which might have been somewhat difficult to eliminate had the heats been made concurrently in several furnaces. It was planned to schedule insofar as possible only one frequently ordered grade of steel. If a variety of grades were made, a direct comparison of data for the two limestone types would have been difficult if the distribution of grade types was not similar for heats made with each type of limestone.

With such a plan it might seem that a direct comparison between statistics such as the average length of time to make the heat for each limestone type could be compared directly. Unfortunately, this approach could be very misleading. Many other factors which are unrelated to limestone type could vary from the set of data for limestone X to the set for limestone Y. Before making such a comparison it would be desirable to adjust for differences of these other unrelated factors from set to set of data. Multiple regression analysis provides an equation for estimating these adjustments.

Table 2 is an illustration of the interference of an important variable, the influence of which is unrelated to the effect of limestone type. The table indicates that 26 heats made with limestone Y averaged .39 hours longer in total time than 29 heats made with limestone X. However, the average scrap charging time of heats made with limestone Y was .50 hours longer than that for heats made with limestone X. Because the length of scrap charging time can importantly affect the total heat time, there is a question as to whether the difference in average total heat time was due only to the difference in limestone type, or whether it was due, at least in part, to the difference in scrap charging time. The kind of limestone charged had no bearing on the time required for charging scrap.

From this it may be seen that a direct comparison of average heat times might be misleading. Because other factors might influence this difference to some extent, it became desirable to estimate the effect of such factors upon heat time. Corrections could then be calculated for the included interfering factors, and could be applied to the heat times in such a way that the resulting difference would more likely be due to the effect of limestone type alone.

Multiple regression analysis is a method for the computation of these corrections. Here again, a limitation occurs because of the small amount of data. Table 2 shows that the data for a total of 55 heats were available, with 29 of these heats made with type X limestone. The difficulty of experimentation is illustrated by the availability of data for only 26 comparable heats made with type Y limestone. Such amounts of data ordinarily are regarded as very small for the purpose of multiple regression analysis when several independent variables are to be included. In this case, this application is somewhat different from those usually encountered. There is no particular interest in establishing accurate estimates of the linear relationship between total heat time and any of the several independent variables. It was not intended to use any such estimating equation in an attempt to make precise estimates of the effect of the included independent variables upon heat time. Rather, the objective was to estimate the contributing effects of certain important variables as reflected in the sample data and then to remove the effects of variations of these independent variables, leaving the effect of the two limestone types remaining. In this way the sample set of data might be considered as a small, finite population, and not as a sample at all. Thus, for the purposes of this study, any relationships found might be considered as exact with reference to the data being evaluated.

It must be acknowledged that the estimates of these relationships might become less reliable if more independent variables were included. Also, only variables with suspected strong effects needed to be included, because it was probably not possible to make accurate estimates of the

TABLE 2

AVERAGE TOTAL HEAT TIME AND AVERAGE SCRAP CHARGING TIME BY LIMESTONE TYPES

	Limestone "X"	Limestone "Y"
Average Elapsed Time - Start Charge to Tap - Hours	13.99	14.38
Average Scrap Charging Time—Hours	4.22	4.72
Number of Heats	29	26

effects of relatively weak variables due to the possibility of misleading results being induced by incidental short-run associations.

SOLUTION OF THE PROBLEM

The selection of independent variables was influenced by prior experience in other studies. The dependent variable was the total heat time from start of charging to tap. The independent variables selected were:

1. Scrap Charging Time.
2. Time from Finish Charge Scrap to Start Charge Hot Metal.
3. Lime Boil Time.
4. Sulphur analysis of heat at start of lime boil.
5. Pounds of Feed Ore Added.

Many other possibilities come to mind but experience indicated that those above had the best chance of indicating strong, essential relationships, if any existed.

With such a small amount of data available it was important to check any multiple regression results for reliability. The usual tests of significance on the size of the regression coefficients were applied, as well as observation of the sizes of R^2 and the standard error of estimate. Still another possibility presented itself. This consisted of separating the data into two subgroups according to limestone type. This was done and a regression equation was computed for each subgroup. The data for heats made with limestone type Y were substituted in the regression equation for the type X limestone heats, and vice versa. The estimated heat times then were compared with the actual times and the results examined for consistency.

The trends of actual heat times were followed in general by the estimated values, but the estimated times tended to be on different levels from the actual times. Thus, for example, the average of 29 estimated heat times for the type X limestone heats was 13.99 hours, when estimated from the equation computed with type X data. That this was equal to the actual average should not be surprising since the equation constant is calculated to do just that. However, when the data for the 26 heats from type Y limestone were substituted in the type X equation, the estimates averaged 14.60 hours, as compared to the 14.38 hours which the heats actually averaged. Under the assumption that equation X reflected the performance of limestone X, it estimated that if limestone X had been charged under the conditions existing for the type Y limestone, type X heats might be expected to take .22 hours longer than the type Y heats actually took. The next question was: "Is this difference truly significant?"

An answer to this question could be approximated under the following assumptions: (1) An equation existed which provided valid estimates of

heat times for type X heats with a standard error of estimate of .891 hours; (2) differences of actual minus these estimated heat times were normally distributed, with a standard deviation of .891 hours; (3) these 26 type Y heats randomly represent the process which produced them. Under these three assumptions, is it likely that the 26 type Y heats would average .22 hours lower than estimated by the equation due to chance alone? The foregoing is no more than the ordinary "t-test" for significance of difference of the average of a sample hypothetically drawn from a known population. In this case, t=1.26, while the 5% level of t=2.09. Since the t-value computed was less than the 5% level, it was concluded that the average time for such a set of 26 heats could easily vary that much from the estimated value due to chance causes, and there was, therefore, reason to discount the importance of such a difference.

On the other hand, using the regression equation developed for the 26 type Y heats, the average estimate of heat time for 29 type X heats was 13.30 hours. The actual average heat time for the 29 type X heats was 13.99 hours. Thus, under the assumption that the type Y equation successfully estimated the performance of type Y heats, the difference of .69 hours computed as above indicated that the type X heats actually took a somewhat longer time than would be estimated for type Y heats. The significance of this difference can be tested as above. The standard error of estimate for the Y equation was .860 hours, and t=4.31 as compared with a 1% level of t=2.81. It was concluded that the type Y equation indicated that the type X heat times were significantly longer than would be estimated from the same equation for type Y heats. The cause for this might be the type of limestone, or it might be some associated factor not included in the equation.

The third test consisted of combining the data and finding the equation for the entire 55 heats. When this was done the estimates for type X and type Y heats were again examined. The average estimated heat time was 13.76 hours for the 29 type X heats compared with 13.99 hours actual. The average estimated heat time for the 26 type Y heats was 14.64 hours compared with 14.38 hours actual. With the combined equation, type X heats were .23 hours <u>longer</u> on the average than estimated, while type Y heats were .26 hours <u>shorter</u> on the average than estimated. The total <u>difference</u> between heats of the two types was .49 hours on the average, after variations of the five independent variables were compensated by the regression equation. This residual difference could also be tested, using some appropriate assumptions.

If it were assumed that heats made with each type of limestone averaged the <u>same</u> time, then the regression equation might be expected to estimate heat times with a standard error of estimate of .903 hours. Assuming normality, etc., as before, the question is:"Could a difference as large as .49 hours be expected by chance between the averages of 29 and 26 heats when sampled from a population of differences having a standard deviation of .903 hours?" This is a common significance test. The only -to-be-expected difference was computed and found to be .244 hours and the associated t=2.01 (the 5% level t=2.01). It was, therefore, concluded that there might be <u>some</u> evidence that this difference was not due to chance causes alone. However, one time in twenty such a conclusion could easily be wrong.

The foregoing values may be more readily viewed and compared by referring to the tabulation shown in Table 3. It was concluded from the foregoing analyses that there was a reasonable indication that the heat

TABLE 3

USE OF REGRESSION EQUATIONS TO CHECK CONSISTENCY OF RESULTS OF ANALYSES

Type X Data Equation	Estimated Avg. Heat Time	Actual Avg. Heat Time	Diff.	S_y	t	t.05	t.01
Type X Heats	13.99	13.99	-	0.891	-	-	-
Type Y Heats	14.60	14.38	- .22	0.891	-1.26	2.09	2.85
Type Y Data Equation							
Type X Heats	13.30	13.99	+ 0.69	0.860	+ 4.31	2.07	2.81
Type Y Heats	14.38	14.38	-	0.860	-	-	-
Combined Data Equation							
Type X Heats	13.76	13.99	+ .23	0.903			
Type Y Heats	14.64	14.38	-.26	0.903			
Difference (X - Y)			.49		2.01	2.01	2.68

times were probably affected by the limestones used. Without presenting the detail here, it can be stated that much other statistical work was performed on the data. Many factors which might influence heat times were tested by significance tests to determine if they were significantly higher for the type X or type Y heats. Since very little evidence of this nature was found it was concluded that heats made with type X limestone would probably have longer heat times than heats made with type Y limestone, all other things being equal. In view of the assumptions which had to be made no attempt was made to put confidence limits on the magnitude of this difference.

CONCLUSIONS

Linear regression analysis, both simple and multiple, is a valuable statistical tool having wide application to the study of steel plant problems. Although the textbooks provide the underlying philosophy and the necessary mathematics, each application will be found to be a distinctive case and a problem in itself. Even when the amount of available data are small the technique may provide valuable and highly rewarding estimates of process complexities.

Considerable care and planning must precede the actual application. Familiarity with the process under study is essential. Interpretations must be cautiously advanced and must include not only the statistical but the practical side. Possibility of the presence of curvilinear relationships, failure to include important variables, change of conditions with time, etc., must be recognized. It is a highly practical tool if used with discretion.

Reference 1. - L. R. Salvosa, "Tables of Pearson's Type III Function," Annals of Mathematical Statistics, Vol. I, No. 2, May, 1930.

HOW TO USE $\bar{X}$ AND R CHARTS EFFECTIVELY

Harmon S. Bayer
Quality Control Consultant

The $\bar{X}$ and R Chart technique has hosts of enthusiastic supporters, including the writer, who can give countless examples of its successes. However, there are many who claim that the technique is overrated. The author finds that when he analyzes any particular situation where a $\bar{X}$ and R chart has not been effective, it almost invariably reflects a failure to intelligently apply the chart to the situation. The purpose of this paper is to illustrate some principles which should be observed when using a $\bar{X}$ and R chart and to outline some of the pitfalls that can be avoided.

The discussion will not include those charts which fail to be useful because production, engineering and other departments do not take action. The case in point is the chart which gets action but still seems ineffective in solving the quality problem.

The main cause of failure of a $\bar{X}$ and R chart to be effective can be summed up as follows: Usually the quality control engineer has not used a chart which is intelligently associated with the physical and engineering facets of the quality problem. A chart will not be effective unless it rings a bell in the mind of the job setter of the equipment involved. It is reasonable to assume that the job setter has little or no knowledge of statistics. If the chart is to help him, the ups and downs of the lines on the paper must in some way reflect not only the variation in the dimension involved, but, in addition, it must reflect or suggest a physical motion which he can associate with his adjustment procedures on the machine. If the chart is not designed to easily establish these associations, in all probability it will be so much wallpaper.

* * *

In order to illustrate these principles let us go through a series of examples beginning with a straightforward problem - one that poses little difficulty in the design of the chart.

Case No. 1 - A straightforward application

Situation: A counterbore and face operation was carried out on a drill press. The characteristic to be controlled was the depth of the counterbore.

History: Customer complaints on this part had been constant. Scrap, rework, and inspection costs were unusually high. The shop personnel had found it extremely difficult to control this counterbore dimension.

Design of the Chart: The quality control engineer applied a simple $\bar{X}$ and R chart (see Fig. A). The variation in the average of a sample of 5 pieces checked for counterbore measurements was indicated on the $\bar{X}$. The difference between the highest and the lowest reading within each sample of 5 pieces was indicated by the R chart.

Analysis: The chart exhibited a range in control, but the $\bar{X}$ showed

a badly out of control situation. This indicated that the process was capable of being controlled.

Action: The quality control engineer contacted the job setter, explaining the chart to him. They worked together for about a day, at which time the job setter found that the stop device on the spindle was not strong enough to control the dimension. The maintenance department was requested to rebuild the stop device.

Result: An immediate improvement in the operation was noted. The $\bar{X}$ chart showed good control after this action (Fig. A). The scrap and rework dropped to less than 0.1%; the excess inspection was removed and the customer complaints on this characteristic ceased.

Discussion: Note the significant fact that the range (R) was in control while the average ($\bar{X}$) was out of control. This usually means a machine capable of being much more consistent than the $\bar{X}$ chart indicates. In many cases simple adjustments of tooling or fixtures can correct this.

Case No. 2 - Another straightforward application

Situation: A drill press was used to drill and ream a hole. The ream dimension was under study.

History: Scrap, rework, customer complaints and inspection costs were excessive.

Design of the chart: Again the $\bar{X}$ and R chart was a simple one (see Fig. B). The variation in the average ream dimension of samples of 5 pieces was shown on the $\bar{X}$ chart. The R chart indicated the difference between the highest and lowest reading in a sample of 5 pieces.

Analysis: When the chart was placed at the job it was readily apparent that both the average ($\bar{X}$) and the range (R) were not in statistical control. This indicated an inability to produce the part with any degree of consistency.

Action: A conference was held with the production, maintenance and engineering personnel. The committee observed the operation and recommended that the fixture be redesigned.

Result: Both the $\bar{X}$ and R chart exhibited control after the redesigned fixture was placed on the operation. The scrap, rework, and excessive inspection costs were considerably minimized and the customer complaints ceased.

So far so good! Nothing difficult here. True! But unfortunately, as many can attest it is not always just that simple. Let's consider one a little more difficult.

Case No. 3 - An out of parallelism problem

Situation: Parallel faces were being cut on a flange by a former on a hand screw machine (see Fig. C). The specification was that the K dimension at any spot perpendicular to the faces should be parallel within a stated tolerance.

History: The usual high cost of scrap, rework, and inspection were present. In this case the customer complaints were unusually serious.

The first chart: A $\bar{X}$ and R chart was applied as shown in Figure D. This was the same type that has been illustrated in Cases No. 1 and No. 2. The chart in this situation however showed no indication of improvement within a reasonable length of time. Conferences were held with the job setter who stated that he was unable to make sense out of the chart. The engineer studied this job further and reached the conclusion that he had failed to adequately analyze the situation.

The second chart: The engineer decided that a single line which only indicated the K measurement taken at random on the part was not sufficiently characteristic of the problem. He concluded that he had to measure each part; locating the maximum K dimension and the minimum K dimension by rotating the part in the indicating gage. He then decided to use two lines on the $\bar{X}$ chart and two lines on the R chart. The top line on the $\bar{X}$ chart would indicate the average ($\bar{X}_{Max}$) of a sample of 5 pieces measured for the maximum K dimension and the lower line would indicate the average ($\bar{X}_{Min}$) of a sample of the same 5 pieces measured for the minimum K dimension. The R chart would also contain two lines, one for the variation within a sample of 5 maximum K dimensions (R_{Max}) and the other for the variation within a sample of 5 minimum K dimensions (R_{Min}).

When the second chart was placed on the machine, the job setter concluded that the out of parallelism on the piece was excessive and that probably the chuck was out of square. This proved to be the case. He had the face of the chuck reground.

The maximum and minimum lines on the $\bar{X}$ chart then showed a decidedly different picture. The extremes were much closer together (see Fig. E). But the maximum line ($\bar{X}_{Max}$) continued to show out of control points. A further discussion with the job setter revealed that the two lines on the $\bar{X}$ chart had helped him realize that the out of parallelism was the problem but now they tended to confuse him when he tried to adjust his average reading. This prompted the development of the third chart (Fig. F).

The third chart: It was reasoned that if the job setter were given target lines for both maximum and minimum $\bar{X}$'s it might be easier to adjust his machine. He was used to a $\bar{X}$ and R chart with a single $\bar{X}$ line to attempt to adjust between control limits and to shoot for an overall average ($\bar{\bar{X}}$). The double line was something different. Since the difference between the maximum and minimum $\bar{X}$'s showed a consistent pattern, the average difference was determined and divided by two. This amount was measured plus and minus from the overall average ($\bar{\bar{X}}$) and at these points on the chart target lines were placed for the maximum and minimum $\bar{X}$ dimensions. Please note Fig. F. To avoid confusion, the upper control limit for the minimum dimension and the lower control limit for the maximum dimension were excluded from the $\bar{X}$ chart. It seemed logical to only include the upper control limit for the maximum (UCL_{Max}) and the lower control limit for the minimum (LCL_{Min}).

Result: The third chart was placed on the job and the job setter now found he was able to use the new chart to aid him in adjusting his averages. The chart reflected good control of the operation within the desired limits. Subsequent to this action, considerable improvement was

noted in the scrap, rework, inspection labor and customer complaint areas.

Case No. 4 - Direction as a significant factor

Situation: A multiple spindle chucking machine performed a number of operations on the piece. The quality problem centered around a hole that was drilled and reamed. The blueprint called for the reamed hole to be centered within a specified tolerance between two parallel edges (a and b) of the piece (see Fig. G).

Let us pause a moment and state that the method of attack toward multiple spindle problems can be the subject of an entire paper. Since this is not primarily the purpose of this paper, we will just make the statement that the data from each spindle has to be handled separately. The following account will cover the method for only one spindle; all others were handled in a similar manner.

History: The usual history of excessive costs in scrap, rework, and inspection labor was present. The customer complaint situation was more than serious.

The first chart. A chart was applied to the job but there was no success in controlling the operation (see Fig. J).

A special study was conducted which indicated that another factor came into play. In the first chart (Fig. J) only the magnitude of the off center dimension was recorded. The result of a check of 100 pieces was analyzed by a histogram and is shown in Fig. L. Note that the specification is shown as 0-6 thousandths of an inch.

A further study of the chucking problem indicated that the direction in which the piece was chucked was extremely important (see Fig. H). It was decided that a manufacturer's identification mark (P) would be considered as a point of relationship to which the off center measurement of each piece would be related. When the piece was placed in the chuck, the face nearest to the P mark was identified and marked. This face was always placed in the same direction in the indicating gage which was set so that "0" would indicate a perfectly centered piece. If the center of the hole was off center in the direction of the marked face of a part (nearest to the P mark on the chuck), the piece was said to have a positive (+) off center measurement. If the center of the hole was off center in the direction of the face opposite the P mark, the piece was said to have a negative (-) off center measurement. Of course, this could be read directly from the indicating gage dial. 100 pieces were then checked taking this direction into account. The histogram of these dimensions is shown in Fig. M.

It can be seen that when we dealt with the tolerance originally, concerning ourselves with only the magnitude, our tolerance was only 6 thousandths of an inch. By taking direction into account we now have a 12 thousandths tolerance (± 6) resulting from +6 in the direction of the P mark and -6 in the direction opposite the P mark. Even more significant is that this is not merely a synthetic assignment of plus and minus symbols. It actually reflects a magnitude and direction of a necessary physical adjustment to the center of the chuck in relation to the P mark of the chuck.

If we refer to Figure M, it is now apparent that the pieces are being centered by the chuck at about 3 thousandths of an inch toward the P mark. Also we note that we now show a fairly normal distribution whereas the distribution without regard to direction is skewed (see Fig. L).

The second chart: A new chart with the $\bar{X}$ scale showing plus and minus readings for the off center dimension was put on the job and the significance explained to the job setter. Note that the $\bar{X}$ chart also indicated that the pieces were centered in the chuck at about 3 thousandths toward the plus side which was nearest the P mark. It was readily apparent to the job center that he had to readjust his chuck centering and the result is shown on the $\bar{X}$ and R chart (Fig. K) which indicates that the job is now well in control.

Result: The improvement noted in Figure M continued and the scrap, rework, inspection labor costs and customer complaints were markedly minimized.

* * *

Reviewing these four cases it is apparent that the following criteria must be present if a $\bar{X}$ and R chart is to aid in the solving of a quality problem.

1. The characteristic chosen to be measured must adequately reflect the quality problem. As an example, in Case No. 3 the original dimension which merely measured the width on a random spot on the piece did not reflect the real problem which was the out of parallelism within the width dimension of the piece. In Cases No. 1 and No. 2 this was no problem. The variation of the measurement plotted on the chart reflected the quality problem as is indicated by the rapidity in which the action was taken by the job setter.

2. An intelligent analysis of the physical and engineering aspects of the problem must be made. In Case No. 3 the first step in solving the problem was to realize that the distribution of out of parallelism within the pieces was too great to allow a reasonable amount of piece to piece variation and still stay within the blueprint tolerances. All subsequent steps depended upon this analysis.

In Case No. 4, the concept of the direction of the off center dimension was necessary to the solution. In addition, a considerable knowledge of the machine, its component parts and adjustment characteristics was required. The solving of the problem was assured when all these factors were related by the device of assigning the directional symbols (plus and minus) to the off center measurement in relation to the direction in which the piece was placed in the chuck.

3. The illustrative methods used on the chart must reflect or suggest to a machine adjuster a physical motion which he can associate with an adjustment or repair procedure on his machine.

In Cases No. 1 and No. 2 this again did not pose a problem. The illustrative devices used on the charts were easily associated with corrective measures in the job setter's mind and he was able to make his corrections.

In Cases No. 3 and No. 4 this was not so simply solved. In Case No. 3 until the maximum and minimum line device was used, no relationship between the chart and the problem on the machine could be established by the job setter. Upon seeing the spread of the two extremes in an easily understandable fashion (the two lines) and comparing them with the specifications he quickly associated the lines on the chart with the off-squareness in the chucking device and corrected that situation. However he was still unable to completely correct his job. The third chart with the two target lines for the average maximum and average minimum dimensions helped him form the additional association with the adjustment characteristics which aided him in correcting the overall average.

Case No. 4 is cited as a classic example of how the illustrative method must reflect a physical adjustment. Before the direction of the off center dimension was determined, the chart only indicated that something was wrong but gave no clue to the operator as to the indicated action. The operator was unable to associate his actions in adjusting the machine with the chart until the ups and downs of the line on the chart could be associated with the direction in which he had to adjust his chuck.

4. <u>It is well to remember that charts cannot substitute for know-how; they can only supplement know-how.</u> Note that the charts did not solve the situation. It was action by people that made the situation right so that the chart was able to reflect an improved situation. This fact is not recognized by some who seem to think that the chart is the major factor which corrects the problem. Make no mistake; nicely designed charts from the standpoint of technical criteria are much more effective when <u>willing cooperation</u> is given by production, engineering, maintenance and other departments.

5. <u>Note that each of these problems was worked out by an aggressive quality control engineer who got out in the shop and dug into the problem.</u> Unfortunately many engineers do not realize that these situations cannot be solved by sitting at a desk in the office. It is a must that to be a good quality control engineer one must become highly trained in the know-how of any particular situation with which he is dealing. This usually means getting out into the shop, getting his hands dirty, studying blueprints, analyzing machine characteristics, and in general really exposing himself to all facets of the problem.

<u>Conclusions</u>: A $\bar{X}$ and R chart is a highly effective tool that can be used in an infinite number of difficult problems. However it must be intelligently applied with due regard for sound engineering concepts and practices. Of equal importance is an ability to use adequate illustrative devices which shop personnel can understand. Failure of a particular chart to be useful can usually be attributed to a failure to observe these principles. Conversely those who use the technique with a reasonable amount of ingenuity, sound methods and some psychological intuitiveness in serving the needs of the shop personnel find the results highly gratifying.

FIG. A

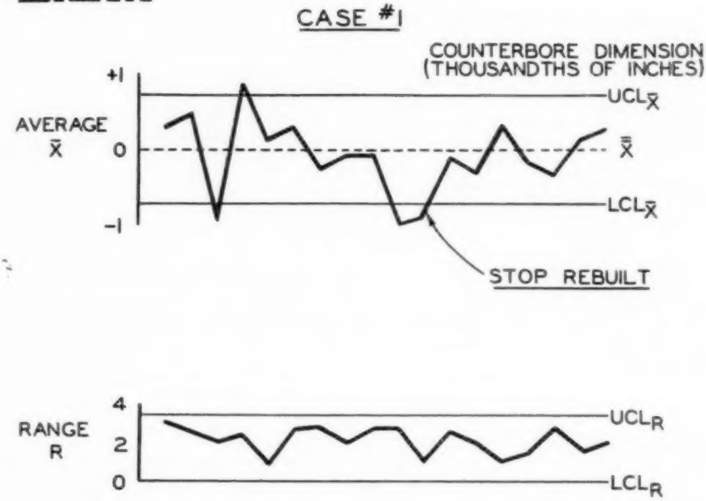

CASE #1
COUNTERBORE DIMENSION
(THOUSANDTHS OF INCHES)
+1
AVERAGE
$\bar{X}$
0
-1
$UCL_{\bar{X}}$
$\bar{\bar{X}}$
$LCL_{\bar{X}}$
STOP REBUILT
RANGE
R
4
2
0
UCL_R
LCL_R

FIG. B

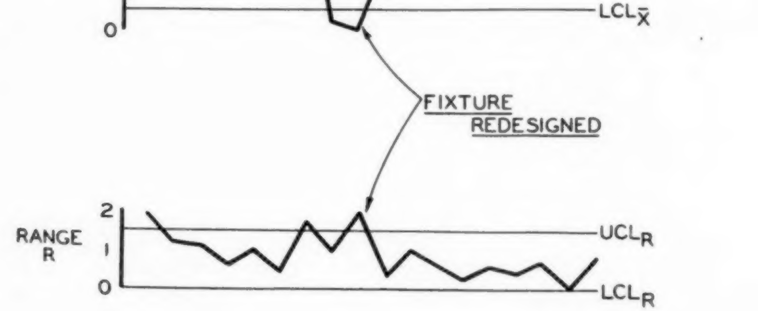

CASE #2
REAM DIMENSION
(THOUSANDTHS OF INCHES)
2
AVERAGE
$\bar{X}$
1
0
$UCL_{\bar{X}}$
$\bar{\bar{X}}$
$LCL_{\bar{X}}$
FIXTURE
REDESIGNED
RANGE
R
2
1
0
UCL_R
LCL_R

FIG. D

FIG. C

CASE #3
THE FIRST CHART

CASE #3

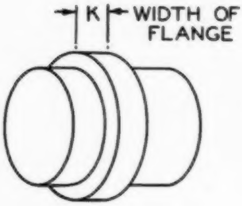


ILLUSTRATION OF SHAPE OF PART

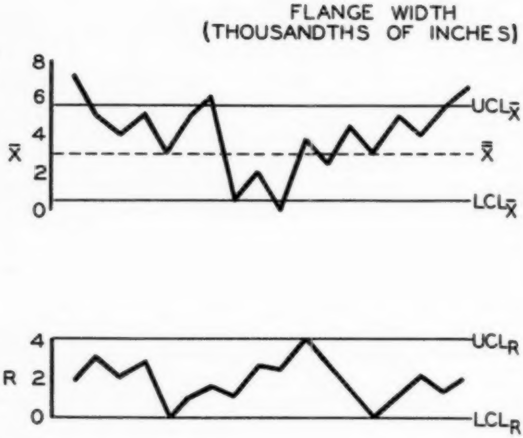


FIG. E

CASE #3
THE SECOND CHART

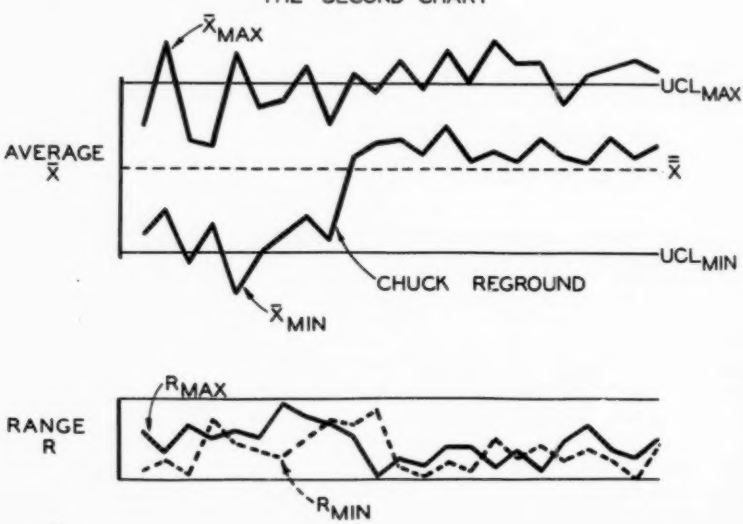


FIG. F

CASE # 3

THE THIRD CHART

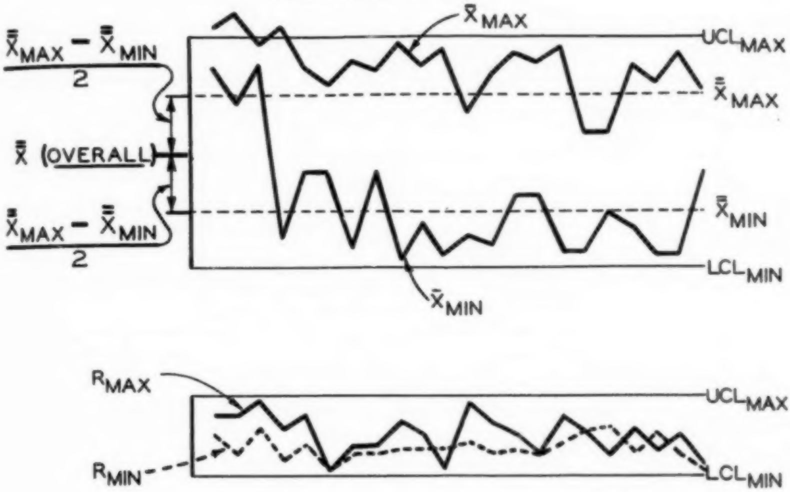
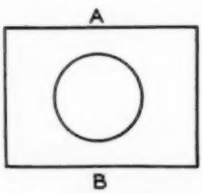


FIG. G

CASE #4

ILLUSTRATION OF SHAPE OF PART



BLUE PRINT CALLED FOR REAMED HOLE TO BE CENTERED BETWEEN FACES "A" AND "B" WITHIN SPECIFIED TOLERANCES.

FIG. H

CASE #4

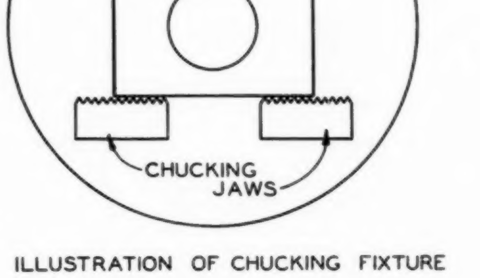


ILLUSTRATION OF CHUCKING FIXTURE

FIG. J

CASE #4
THE FIRST CHART

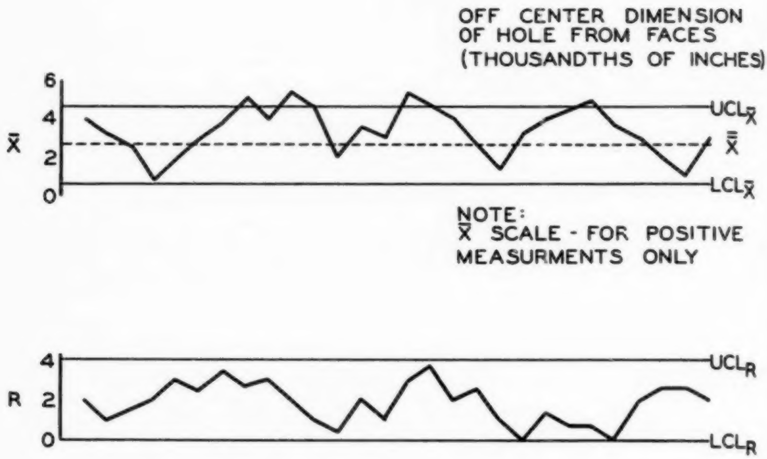


FIG. K

CASE #4
THE SECOND CHART

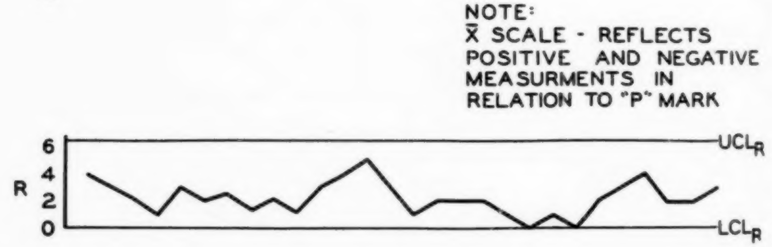
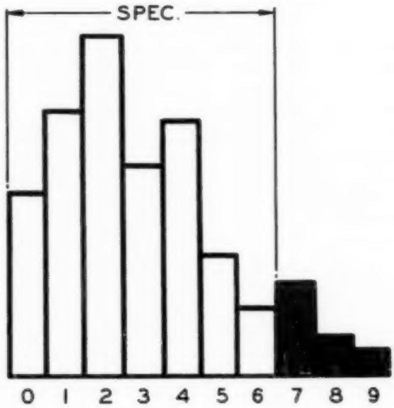


FIG. L

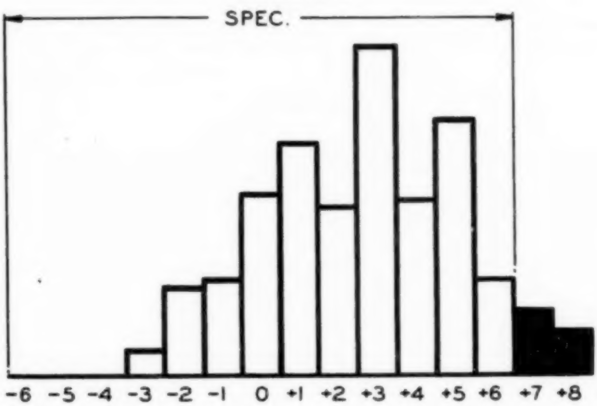
CASE #4



HISTOGRAM OF VALUES OF OFF CENTER DIMENSION
DIRECTION IGNORED
POSITIVE VALUES ONLY

FIG. M

CASE #4



HISTOGRAM OF VALUES OF OFF CENTER DIMENSION
DIRECTION IN RELATION TO "P" MARK ON CHUCK
POSITIVE AND NEGATIVE VALUES

VARIABLES THEORY - FREQUENCY DISTRIBUTIONS

Robert W. Boeke
John Deere Ottumwa Works
Deere Manufacturing Company

In January, 1798, Eli Whitney obtained a contract to supply the United States Government with 10,000 muskets. It was his aim "to make the same parts of different guns, as the locks, for example, as much like each other as the successive impressions of a copper plate engraving." The fulfillment of this contract marked the first successful application of the principle of interchangeable manufacture. Furthermore, it attests to a technological advancement which permitted the restriction of the quality of component parts within such narrow limits that randomly selected components could be assembled into an adequately functioning mechanism. Little, if anything, was known regarding the variation which existed among similar components other than that it was sufficiently small to permit interchange. It remained for Dr. W. A. Shewhart, in the early 1920's, to provide a procedure for analyzing and controlling the variation. Statistical methods provide the tools for economically controlling quality within the confines dictated by interchangeable manufacture.

Eli Whitney is credited with pioneering the principle of <u>Restrictive Quality</u> through the introduction of Interchangeable Manufacture; Dr. Shewhart, for the principle of <u>Controlling Quality</u> through Statistical Quality Control. Modern day requirements for economical and efficient manufacture demand utilization of both these principles to attain maximum productivity. Effective use of statistical methods requires a knowledge of fundamental theory.

The purpose of this paper is to outline, briefly, the theory which underlies the statistical methods which are applied to measurement (variables) data. The theory will be developed by an intuitive rather than a mathematical approach.

INTRODUCTION

For any given characteristic, measurements of adequate precision will reveal variations among "similar" parts, batches, or runs. This variation appears inevitable in nature and any attempts to remove it are futile. It can be reduced, but never eliminated.

If a large number of measurements are made on similar parts, it is possible to illustrate the variation by means of graphical methods and thus determine the pattern of variation for that particular characteristic. If the manufacturing conditions, procedures, and raw material remain stable, it can be inferred that this pattern of variation will be duplicated within close limits, by subsequent measurements. This procedure of making estimations or inferences concerning a universe, or population, based upon limited data is the core of statistical methods.

ILLUSTRATIVE PROBLEM

As an example, let us suppose that we are asked to determine the tensile strength characteristics of 30,000, 5/8-inch hardened steel bolts. Although it would be possible to test each of the bolts, it would not be feasible since the test is a destructive one. Our problem,

then, is to make estimates regarding the entire universe, which in this case is 30,000 bolts, based upon measurements obtained from a sample. It is evident that we must be careful to select a sample which is representative of the entire universe. This can most often be obtained by selecting the sample items at random from the entire lot. The size of sample necessary depends upon the reliability desired in estimating the universe. The larger the sample, the greater the reliability of the results. We will assume that a sample of 50 bolts was tested and that the data in Table I were obtained.

TABLE I

Tensile Strength of 50 5/8-inch Steel Bolts Selected at Random from 30,000.

29,950	31,500	30,500	32,500	33,000
35,600	34,150	32,400	30,350	28,500
28,900	29,150	31,100	31,300	33,100
31,300	33,350	31,100	32,700	31,400
32,150	33,450	31,900	31,300	31,150
32,300	31,750	32,250	32,000	32,500
30,900	31,600	30,400	32,750	31,500
33,000	29,800	30,100	30,500	32,800
29,600	27,200	34,950	31,050	31,050
30,800	29,000	32,150	31,100	36,300

The data in Table I are difficult to comprehend since there is no order or arrangement. It is nearly impossible to extract any pertinent information with data in this form; consequently, a table, such as the frequency distribution in Table II, becomes valuable.

TABLE II

Frequency Distribution of Tensile Strength of 50 5/8-inch Steel Bolts Selected at Random from 30,000

Class Midpoint	Frequency
27500	1
28500	2
29500	5
30500	7
31500	15
32500	11
33500	5
34500	2
35500	1
36500	1

Although the data in Table II has lost some of the detail of Table I, this loss has been more than offset by the gain in simplicity. Actually, the loss of detail is often greatly over-emphasized. If another random sample of 50 bolts were selected, it is almost certain that their measurements would not correspond exactly to those in Table I. Furthermore, there is no reason to prefer one sample over the other and, consequently, the detail is only of secondary importance. The sample is to be used to make estimates concerning the entire group from which the sample is drawn. The factors of utmost importance are the measures of central tendency (average) and dispersion (standard deviation) and the shape of the distribution, which will indicate the proportion of parts of various sizes. The loss of the individual identity of the original measurements is not important; as a matter of fact, it is highly important to think of the data in terms of intervals, rather than individual measurements.

To further simplify comprehension of the data, it may be shown graphically as a histogram. See Figure I.

The histogram in Figure I is characteristic of many histograms which are obtained in industry. Before we attempt to make any estimates concerning the 30,000 bolts (universe) from which the sample of 50 was

drawn, let us analyze critically the procedure we've followed. A random sample was drawn, measurements of tensile strength were made, and the data were classified in cells, or intervals, of 1000. It is not too difficult to acknowledge that if another sample of 50 were to be measured that the second 50 measurements would not correspond exactly to the original 50. Furthermore, it is highly improbable that a second histogram would be the same, in all respects, as Figure I. It is unlikely that we would get exactly 15 measurements between 31,000 and 32,000 pounds, or exactly 7 between 30,000 and 31,000, etc.

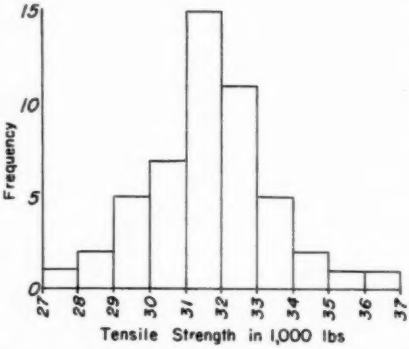


Figure I- Histogram of Tensile Strength of 50 5/8-inch Steel Bolts Selected at Random from 30,000.

ESTIMATING THE UNIVERSE

If the data cannot be precisely duplicated, is it of any use? Although other histograms would differ in detail from what has been obtained, there are certain characteristics which are reproducible within very close limits. These reproducible qualities are:

1. The average (central tendency)
2. The standard deviation (dispersion)
3. The pattern of variation (shape)

Although other measures of central tendency such as the mode and median are occasionally useful, the most common measure is the arithmetic mean. The mean is calculated by dividing the sum of all measurements by the number of measurements. For the illustrative problem, this value was calculated to be 31,583 pounds.

The root-mean-square deviation of the individual measurements about the mean provides a measure of dispersion, the standard deviation. For the illustrative problem, the standard deviation is 1,727 pounds.

The pattern of variation, illustrated in Figure I, has been described as not exactly reproducible. However, there are certain characteristics which are stable.

Figure I shows that the large majority of the measurements are near the mean, with only a small proportion near the extreme limits. Even though many more samples of 50 were drawn, it is inconceivable that this ratio would change appreciably. Each histogram would be expected to contain a large proportion of the measurements near the mean, with decreasing frequencies as the extreme limits were approached. This attribute is characteristic of most, not all, measurement data.

Since the sampling was performed to draw inferences about the entire 30,000 bolts, there is no reason to prefer any one histogram of 50 measurements over any other. It may be assumed that the measurements of the universe would follow a pattern which could best be described by a smooth curve which possesses the characteristic shape of measurement

data. This model, or generalized histogram, is called the Normal Curve.

The Normal Curve, shown in Figure II, is a bilaterally symmetrical curve with a maximum ordinate at the mean. 34.13% of the area under the curve is contained in the first standard deviation on either side of the mean, 13.59% in the second standard deviation, and 2.14% in the third.

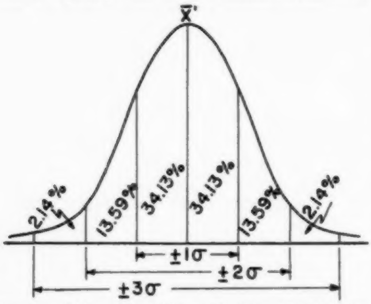


Figure II - The Normal Curve Showing the Respective Areas Contained in Various Standard Deviation Zones on Either Side of the Mean.

Since the area under the Normal Curve, as with the histogram, is related to the frequencies of occurrence, it may be said that approximately 68.27% of the measurements will lie within one standard deviation of the mean, 95.45% within two standard deviations of the mean, and 99.73% within three standard deviations. Additional frequencies can be determined from a "Table of Areas Under the Normal Curve".

The data from the sample of 50 tensile strength measurements have been condensed to two statistics and an assumption regarding the shape of the parent distribution. The sample average is used to estimate the average of the universe, the sample standard deviation to estimate the standard deviation of the universe, and the assumption of the Normal Curve is used to describe the pattern of variation. See Figure III.

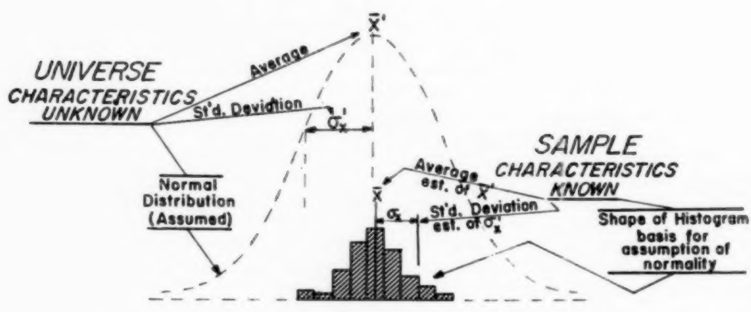


Figure III - Estimation of Universe Characteristics Based Upon Sample Information.

Since estimates have been made of the average, standard deviation, and shape of the distribution of the tensile strengths of the 30,000 bolts, it is possible to make some estimates regarding the percentage of bolts which lie in various intervals. Thus, the proportion of bolts with a tensile strength of less than 29,000 pounds may be estimated; or the proportion between various other limits can also be determined. This technique, using the Table of Areas Under The Normal Curve, is particularly useful when comparing a set of data with specifications, design limits, etc., or in establishing process capabilities. As an

illustration of this procedure, reference should be made to Figure IV where the calculation of the percentage of bolts with a tensile strength less than 29,000 pounds has been shown in diagram form. It should be noted that the illustration represents the distribution of the entire 30,000 bolts and is described in terms of the sample average, sample standard deviation, and the Normal Curve.

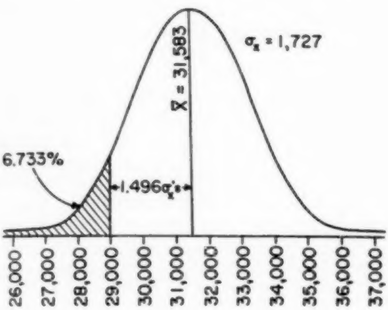


Figure IV - Estimated Distribution of the Tensile Strength of 30,000 5/8-in. Steel Bolts.

The similarity between Figures I and Figure IV should be observed. The histogram in Figure I represents the actual measurements obtained in a sample of 50 bolts; the Normal Curve in Figure IV was derived from the sampling data and is used to represent the 30,000 bolts from which the sample was drawn. This procedure of generalizing from the particular, characterizes statistical techniques.

It will be noticed in Figure IV that 6.733% of the bolts are estimated to have tensile strength less than 29,000 pounds. This percentage value is taken from the "Table of Areas Under The Normal Curve". We might very well question the preciseness of 6.733%. Are we justified in making such an accurate estimate, or does our data only permit a rough approximation of the order of "approximately 6%"?

TESTS OF RELIABILITY

To determine the precision warranted in estimating the percentage of values beyond a certain limit, 6.240% for instance, let us look at the precision of the data utilized in calculating this value. We know that if other random samples, of the same size, had been drawn from the 30,000 bolts that it is highly improbable that we would have obtained precisely the same average, 31,583. We are likely, then, to question the precision with which this value estimates the true population average. What confidence can we place in our sample value (statistic)?

If many samples, each of 50 bolts, had been tested, the sample averages would vary. Furthermore, the magnitude of this variation is a function of the sample size. The larger the sample, the less the variation. The relationship between the variation of the individual measurements and the sample averages may be shown by the following relationship:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

The standard error of the mean is equal to the standard deviation of the individuals divided by the square root of the sample size. For our illustrative problem $\sigma_x = 1727$ and $n = 50$. Thus, $\sigma_{\bar{x}} = 244$.

Using Student's t-Distribution, we may make the following interval

statement regarding the mean of the 30,000 bolts. We are 95% confident that the true population mean lies within the interval of 31,583 ± (2.01)(2.44), or 31,583 ± 490. Although 31,583 is still our best estimate of the population mean, our confidence limits indicate that it may be in error by as much as 490 pounds. To further clarify this measure of uncertainty, it can also be said that we are 50% confident that the true population mean lies in the interval, 31,583 ± 166. Thus the likelihood of being in error by as much as 490 pounds is somewhat remote; it is just as likely that our error is 166 pounds or less, as it is to be 166 pounds or more.

Just as the sample average is subject to sampling fluctuation, so, also, is the sample standard deviation. It can be shown that although the sample standard deviation, 1727, is our best estimate of the population standard deviation, we are only 90% confident that the population standard deviation lies in the interval, 1486 to 2072. For the academic minded, this interval was calculated by using the F-Distribution ---- first $F_{.05}$ ($n_1 = 50$, $n_2 = \infty$) and then $F_{.05}$ ($n_1 = \infty$, $n_2 = 50$). Here again we have been able to obtain a measure of the reliability of our estimates. For large sample sizes this interval may be approached using $\sigma_x \pm 1.645\sigma_\sigma$, where

$$\sigma_\sigma = \frac{\sigma_x}{\sqrt{2n}}$$

An assumption has been made regarding the normal distribution of the measurements. We may wish to test the validity of this assumption. Many different methods have been used to test the normality of measurement data, among which are:

1. The "look" test.
2. The use of probability paper.
3. The χ^2 test for goodness of fit.

Most often, a "look" test, applied to a histogram provides adequate assurance regarding normality. However, for marginal histograms, particularly with small sample sizes, this procedure may lead to erroneous assumptions. It has been shown, by classroom demonstrations, that the judgment of people acquainted with statistical methods will vary considerably regarding the allowable deviations from true symmetry and conformity. This condition indicates the need for a more precise technique.

Probability paper is constructed in such a manner that a cumulative frequency distribution, for a perfectly normal distribution, will plot as a straight line. Since all histograms, and their companion cumulative frequency distributions, exhibit some irregularity, the plottings never lay on a perfectly straight line. The problem then becomes that of determining how much variation can be tolerated and still allow an assumption of normalcy. Basically, then, this procedure also amounts to a "look" test.

The χ^2 test for goodness of fit permits us to determine the probability that the sampling distribution could have been obtained from a normal distribution. If this probability is reasonably large, we may safely assume the normal distribution; if small, it will be necessary to eliminate the assumption of normalcy. In our illustrative problem, the χ^2 test indicates that more than 80% of the samples (n = 50) drawn from a normal distribution would exhibit as much or more irregularity. Since

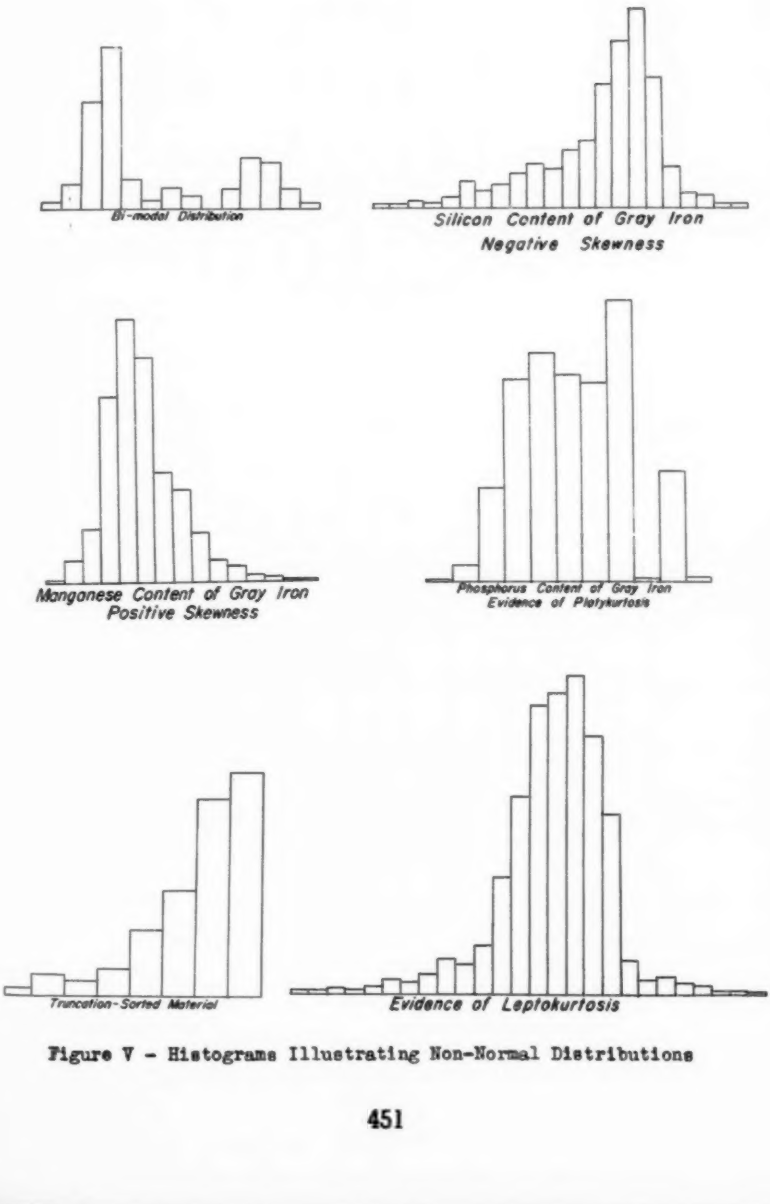


Figure V - Histograms Illustrating Non-Normal Distributions

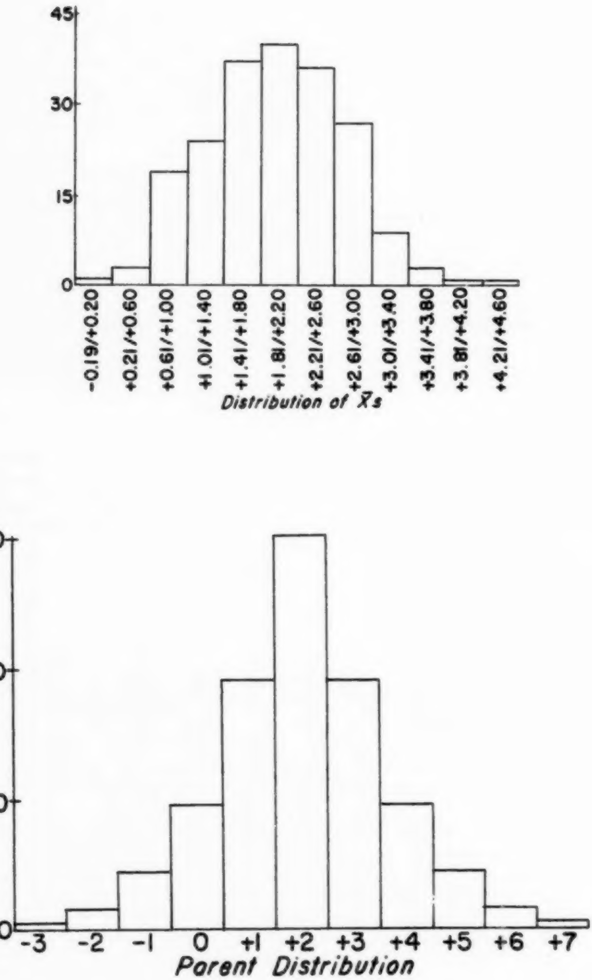
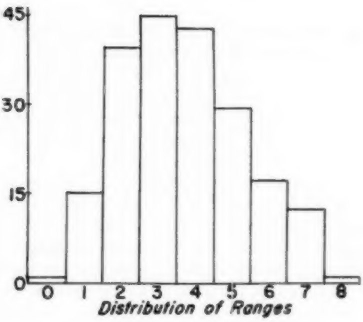


Figure VI - Histograms of the Parent Distribution and Sampling Distributions of $\bar{X}$.



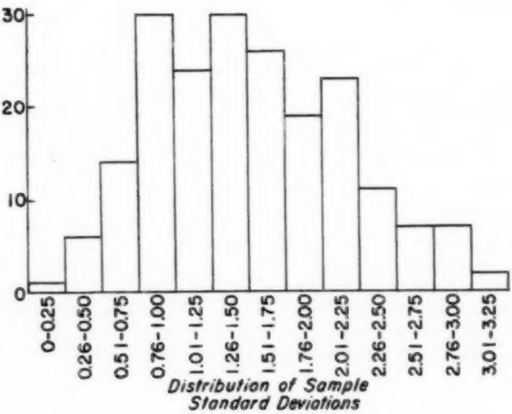


Figure VII - Histograms of Ranges and Standard Deviations From Parent Distribution Shown in Figure VI

this probability is high, we may assume that the tensile strengths of the 30,000 bolts are distributed normally. It should be emphasized that our calculations do not indicate a probability of 80% that the population is normal. (The population either is, or isn't, normal) Our calculations indicate that there is a high probability of obtaining a histogram, similar to the one obtained, from a normal distribution.

PATTERNS OF VARIATION

Unfortunately, all distributions do not follow a normal distribution. Many forms of irregularity occur, such as:

1. Skewness
2. Kurtosis
3. Truncation
4. Bimodality

Illustrations of these forms of irregularity are shown in Figure V. All the illustrations in Figure V are data actually obtained from industrial data. It is rather obvious that an assumption of normalcy could lead to very misleading conclusions. There is no simple procedure for making probability statements regarding non-normal distributions. Although other models similar to the normal curve are available, their usage is limited and requires rather rigorous mathematical treatment.

One method of preparing probability statements utilizes the empirical approach. The relative frequencies occurring in the sample are inferred to exist also in the universe. This procedure, subject to sampling fluctuations, provides very inadequate results unless the sample size is very large.

A very satisfactory technique exists for making probability statements when the distribution is either non-normal or unknown. It states in Tchebycheff's inequality theorem that more than $1-(1/t^2)$ of any set of finite numbers must fall within the closed range, $\bar{X} \pm t\sigma$, for values of t, where t is 1 or larger. Thus, if $t = 3$, at least 8/9ths of the values must lie within the interval $\bar{X} \pm 3\sigma$, where $\bar{X}$ and σ are calculated from the data.

If the data is unimodal and symmetrical, even though not normal, the Camp-Meidell theorem states that more than $1-(1/2.25\ t^2)$ of the values will lie within the limits $\bar{X} \pm t\sigma$. For the above conditions, more than 95.1% of the values will fall within $\bar{X} \pm 3\sigma$. These formulas provide definite probability values for distributions which are non-normal. Although the t-values shown above were for 3σ limits, it should be emphasized that the formulas are valid for all values of t greater than one.

SAMPLING DISTRIBUTIONS

We have mentioned sampling fluctuation, the standard error of means, and Student's t-Distribution. To better understand this terminology, let us resort to an empirical approach. We have set up a hypothetical distribution, based upon the normal curve, which is represented by "measurements" on chips. Random samples of the chips have been selected and the average, standard deviation, and range have been calculated for each. It may be observed that the distribution of individual values was set up according to a near-normal chip bowl with a mean of +2 and limit-

ing values of +7 and -3. The standard deviation of the bowl is 1.64. The distributions of the universe, averages, standard deviations, and ranges are shown in Figure VI and Figure VII.

It will be observed that the parent distribution of individuals, the universe, is approximately normally distributed. The distribution of the sample averages also appears normal with an average of approximately +2.0. The dispersion of the averages, for samples with $n = 5$, is considerably less than the dispersion of the individuals. Theoretically, the standard error of averages, $\sigma_{\bar{x}}$, which is analogous to the standard deviation for individuals, should be equal to $\sigma_x/\sqrt{n}$ or $1.64/\sqrt{5} = 0.733$. This value compares favorably with the value actually obtained from the distribution of averages, 0.736. Although the means appear to be normally distributed, Student's t-Distribution is more precise, particularly when the sample size is small.

The distributions of both σ and R appear to be skewed to the right which is characteristic of these statistics. Although the distribution is not normal, it may be said to follow the χ^2 distribution for small sample sizes, and approaches the normal distribution as the sample size increases. Thus, it is possible to establish confidence limits for standard deviations and ranges. It should be noted that all the sampling distributions are functions of the sample size. As the sample size is increased, the individual statistics become more reliable estimates of the population parameters. An increase in sample size reduces the standard error of estimate for the statistics, and narrows the confidence limits we place on their reliability.

When we understand the patterns of variation and the reliability of sample statistics, we have the knowledge necessary to enhance the application of statistical methods to the control of manufactured products.

HOW TO MAKE DECISIONS

Irwin D. J. Bross
Department of Public Health and Preventive Medicine
Cornell University Medical College

Introduction: We live in a complex, interrelated technological world. We have more knowledge of the world about us; we have more techniques for the control of this world than ever before in the history of man. Our basic problem is to use our new knowledge and powers wisely and efficiently. We must make better decisions if we are to survive in this strange new world of our own making. The vague, intuitive, personal methods for making decisions that were used in the past (and none too successfully even in the simpler decision situations of the past) must give way to more scientific procedures of making decisions.

In this paper I want to discuss very briefly the basic principles that underlie the new scientific procedures for Decision-Making that are currently being devised and developed. Although the procedures themselves are somewhat too technically complex to go into here, the principles are not technical--or even highbrow--they are plain common sense precepts.

I want to try to present these principles in operational form, that is, as a list of steps which can be taken in making practical decisions. Broadly speaking, what we do is to trace down the consequences of each of the alternative lines of action open to us and by comparing the expected consequences, we arrive at a choice of action. This is evidently the common sense way to make a decision and is nothing new. As John Dewey said many years ago: "The true object of knowledge resides in the consequences of directed action." There is a hitch in applying this principle: At the time when the decision must be made we do not know the consequences of directed action. The best that we can do is to predict the consequences (but this prediction must always be made on the basis of incomplete information). In other words, practical decisions must be made in the face of uncertainty and the new Decision-Makers are designed to operate in these circumstances. Thus the new methods of Decision Making are designed to make efficient use of incomplete information and to allow for the incompleteness of the information. Moreover, the procedures are quantitative; they replace the vagueness and ambiguity of verbal principles by the precise and clearly defined language of mathematics and number. The essential novelty of modern Decision Making does not lie in the principles but in the translation of these principles into technological instruments for making decisions.

The Seven Steps to Decision: The great danger in discussing general principles is that they may seem unrelated to practical problems. To avoid this I want to discuss the principles in terms of an example. Although I think this example is fairly realistic, simplicity is even more important for purposes of example and to keep things simple on the technical side, I have sometimes resorted to drastic approximations for which I crave your indulgence.

Step 1. Frame, in general terms, the context of the decision.

Answer such questions as: What is to be decided? What is the underlying situation? What information is to provide the basis for the decision? What are the overall objectives of the decision? Do this in

a general way; the later steps will deal with these points in more detail.

Here is the example that I will consider: A plant is manufacturing a high precision item--I will call them "widgets". Each widget is tested as soon as it is finished (I will use the symbol Y to denote the measurement obtained in this test) and if it fails to meet specifications it must be reworked at a cost of $10 (in delay, time, and tools). The process for manufacturing widgets is complex and hard to keep in control. The process can be recalibrated (the recalibration point will be taken as zero) but this also interrupts production and runs to a total cost of $5. The manufacturer's problem is that if he recalibrates each time his profits will be substantially reduced, but if he doesn't, his rework costs will have an even more drastic effect on his profits.

Step 2. List the alternative courses of action.

I will assume that the manufacturer is considering the following three courses of action (there are evidently many other courses open):

A_1 Recalibrate after each widget.
A_2 Never recalibrate.
A_3 Recalibrate after a widget fails to meet specifications.

Step 3. Trace the possible outcomes of each course of action.

If we recalibrate each time we will nearly always have a widget that will meet specifications. If we never recalibrate the situation is somewhat more complicated. Suppose that we start with the process in calibration. The first widget will nearly always be OK but the process will be drifting off calibration so we are not so sure of the second widget. Here we encounter a very common practical situation which may be called a probability event chain. I will use a subscript to denote the order of production of a series of widgets. Thus Y_1 will denote the initial widget immediately after recalibration and Y_2, Y_3, etc. will designate the succeeding widgets. I will use a prime to indicate a widget requiring rework; thus Y_2' is a situation where the second widget requires rework. The possible outcomes, therefore, consist of all the different sequences such as $Y_1\ Y_2\ Y_3'$ ..... . Such a sequence can be regarded as a chain of events and, since we are uncertain as to which sequence will actually occur, we use the term "probability event chain".

We can proceed in the same way for the case where we recalibrate after a widget requires rework. Notice, however, that if we take action A_3, some of the event chains become impossible. For example, the chain $Y_1\ Y_2'\ Y_3'$ is possible if we never recalibrate but cannot occur with action A_3.

Step 4. Determine the probability of each possible outcome.

Even though we are uncertain which event chain will occur it should be clear that there are degrees of uncertainty; some events may be likely and others very unlikely. We measure the chances of an event by means of a Probability Scale. Events which are certain to occur have probability equal to one; impossible events have probability zero, and in general the probability will be some number (i.e. decimal fraction) between zero and one. The determination of the numerical value leads into technical problems but this is such an important step that I want to carry our example a little further even though it becomes a bit technical.

THE DRIFT FROM CALIBRATION

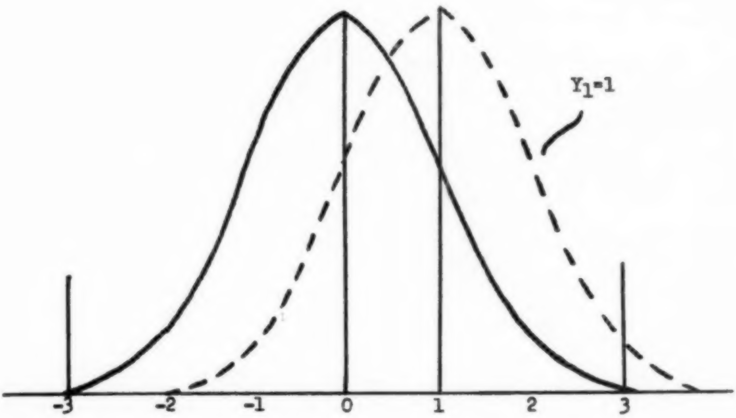


Chart 1a
Distribution of Widgets: After Recalibration

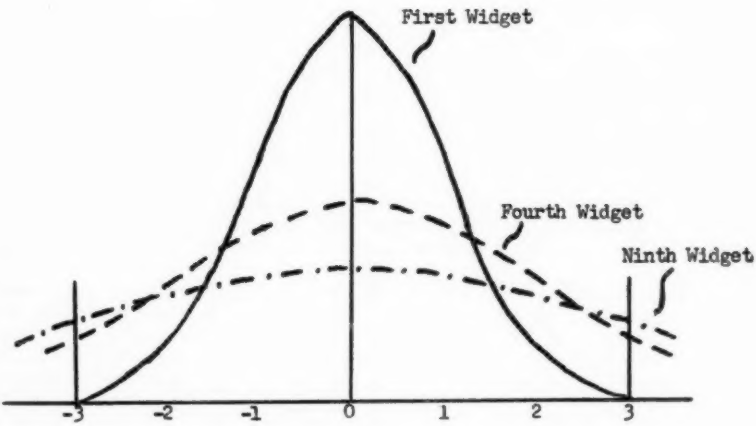


Chart 1b
Distribution of Widgets: No Recalibration

In general there are three ways to calculate probabilities; two of which will be mentioned here. The first is the Direct Method. For example, we might keep a record of what happened when we never recalibrated. The probability that the second widget would require rework would then be estimated by the proportion of second widgets which failed to meet specifications. The Direct Method is often used but it tends to be an expensive way to do the job.

The second approach, the Model Method, is quite different. Here what we do is to set up a mathematical model which we hope will describe our situation. Our estimate is, therefore, the result of what amounts to a theoretical experiment or study rather than an actual study as in the case of the Direct Method. We generally would like to check our theoretical model against reality but much less data is needed to make this check than is needed for the Direct Method.

A Model Describing Drift from Calibration: I will now present a simple model to describe the drifting of the process. Consider first the initial widget (after recalibration). It will not necessarily have a zero measurement; rather most of the widgets will have measurements near zero and large departures from zero will be infrequent (see the solid curve in Chart 1a).

Note: In what follows it will be assumed that the widgets will require rework if their measurement is more than three units above or below zero.

Now consider the second widget. I will assume that it will show the same sort of distribution as the first widget except that its distribution will be centered about the measurement of the first widget. Thus if $Y_1=1$ the second widget would probably have a measurement close to 1 (see the dashed curve in Chart 1a). Note that Y_1 is unlikely to be in the rework zones but if $Y_1=1$ then Y_2 has a much larger chance of falling in the upper rework zone.

This model is called a "random walk" model and is based on the normal distribution. In this example the initial variance (scatter) of the normal distribution is taken as one, but in practice we would estimate the variance from actual data. If this model applies then it is easy to calculate the probabilities for the case when there is no recalibration. The distribution for the nth widget is simply the normal distribution with variance $\sqrt{n}$. These probabilities are pictured in Chart 1b. It is evident that the drift from control is fairly rapid; by the ninth widget there is about one chance in three that rework will be necessary. The numerical values of the probabilities are given in Column 2 of Table 1.

When we recalibrate after a widget requires reworking the probabilities are harder to calculate. We must deal with event chains of different lengths (i.e. the chance that we must calibrate after the first widget, after the second widget, and so forth). Fortunately (after the first few widgets) the chance that the next widget will be defective is about one in ten and this number changes fairly slowly. Hence, we can use a simple formula for the probability that the chain ends on the nth widget:

$$\left\{1.0 - P(Y_1') - P(Y_2') - \dots P(Y_{n-1}')\right\} \times 0.09$$

where $P(Y_1')$ is the probability that the chain ends on the first widget, etc. Note that the quantity in brackets is merely the chance that the

TABLE 1

Calculation of Expected Costs

Col. 1	Col. 2 No Recalibration	Col. 3 Recalibration	Col. 4	Col. 5
Widget Number	Probability that Rework Will Be Necessary	Probability that Chain Ends With n-th Widget	Cost Per Widget if Chain Ends With n-th Widget	Col. 3 x Col. 4
1	.0026	.0026	$15.00	$ .0390
2	.0339	.0339	7.50	.2542
3	.0836	.0867	5.00	.4335
4	.1336	.0789	3.75	.2959
5	.1802	.0718	3.00	.2154
6	.2224	.0653	2.50	.1632
7	.2584	.0594	2.14	.1271
8	.2892	.0541	1.88	.1017
9	.3174	.0492	1.67	.0822
10	.3422	.0448	1.50	.0672
11	.3682	.0408	1.36	.0555
12	.3844	.0371	1.25	.0464
13	.4066	.0338	1.15	.0389
14	.4238	.0308	1.07	.0330
15	.4354	.0280	1.00	.0280
16	.4532	.0255	.94	.0240
17	.4654	.0232	.88	.0204
18	.4776	.0211	.83	.0175
19	.4902	.0192	.79	.0152
20	.5028	.0175	.75	.0131
Total				$ 2.0714

first (n-1) widgets do not fall in the rework zone. Thus the formula follows from common sense. Although Step 4, the calculation of probabilities, is technical, it can quite often be carried out by someone who knows only a few simple rules about combining probability. The technicalities of the procedure used here to obtain the entries in Column 3 of Table 1 are explained in a little more detail in the Appendix.

Step 5. Assess the desirability of each possible outcome.

In this industrial situation it is natural to assess desirabilities in terms of dollars and cents. Hence this step is largely a matter of cost accounting. In this example this step is very easy (but ordinarily it is considerably more difficult).

If we always recalibrate (A_1) the cost per widget will be $5.00. If we never recalibrate (A_2) the cost per widget will depend on the proportion of widgets in our event chain which require rework. If we recalibrate whenever a widget must be reworked (A_3) the cost per widget will be $15.00/n where n is the number of widgets in the chain (see Column 4, Table 1).

Step 6. Calculate the expected consequences of each course of action.

This means that we must combine the probabilities (Step 4) with the desirabilities (Step 5) and here we could do this by calculating the expected cost per widget. In general, the expected cost is the product of a probability and a desirability (or a sum of such products). For example, if we had a process where the probability of a rework was 0.50 then the expected cost of rework would be: 0.50 x $10.00 = $5.00. To calculate the expected cost of A_3 we would use Table 1. We first multiply each number in Column 3 by the corresponding number in Column 4 and then add these products together. Thus, the sum would start out:

$$.0026 \times 15.00 + .0339 \times 7.50 + .0867 \times 5.00 + \ldots$$

In Table 1 the calculations are given for the first 20 widgets and come to $2.07. If the calculations are carried to the 35th widget the value is slightly more ($2.20). Except for very fine decisions, therefore, the first few terms are sufficient for practical problems. The formula given here overestimates expected costs a little (see the Appendix).

If we never recalibrate, the expected costs depend on how long the procedure is continued. Eventually nearly all of the widgets will have to be reworked so that the expected cost per widget tends towards $10.00.

Step 7. Compare the expected consequences and select the most favorable course of action.

All that is necessary in this example is to observe that expected costs for the three courses of action are A_1, $5; A_2, $10; and A_3, $2.20. Note that we are not only led to the selection of A_3 as our course of action but furthermore we can say, in terms of dollars and cents, just how much better off we will be by following this line of action.

Post Mortem: The seven steps to decision listed here should be followed by an eighth step which, while not necessary for this particular decision, will provide a guide to future decisions.

Step 8. Check up on the results.

After we have made our decision it generally pays to follow it up. Did it work out as expected? If not, what went wrong? How can this information be used to improve future decisions? Remember that the proof of the pudding is in the eating. If the results turn out sour this means that something went wrong in our preceding steps. We may have omitted an effective course of action from our list (perhaps we would do better to recalibrate before we wind up in the rework zone). Perhaps our model is haywire. Maybe our costs are unrealistic (perhaps such costs as "employee morale" or "consumer satisfaction" have been neglected). Perhaps instead of dealing with long-run profits (expected costs) we really want to control immediate losses (in which case we might wish to work with a minimax criterion). When we are setting up a fancy new Decision-Maker it generally requires some adjustments. In short, decision making itself is a chain of events and we must continually strive to utilize our past mistakes to improve our future decisions.

Levels of Decision:

It will be obvious to you that major decisions are going to require a more comprehensive treatment than was indicated in my little example. However, though each step becomes more difficult, the same basic principles apply to major decisions. You may have noted that in the example used here there are really three levels of decision involved. A rule such as "recalibrate after a widget requires rework" itself specifies a decision at the shop level. Given this rule, of course, the decision as to whether or not to recalibrate is routine and more or less automatic.

The choice of the rule represents another level of decision, for example, this decision might be the responsibility of a quality control engineer. The choice of rule depends upon decisions at a still higher level. For example, administrative policies would determine whether the goal should be maximizing long term profits or the control of immediate losses.

There is a very important change that takes place as we go from the lower to the higher levels of decision: We tend to shift from routine problems to research problems. The quality control engineer who is trying to construct a model to describe the drift out of calibration is really in much the same sort of situation as a research scientist who is trying to discover the explanation for some natural phenomena. Indeed the tools of a research scientist come into play even in the simple example that we have described.

Major Decisions:

Suppose now that we turn to a higher level decision problem; let us say that we are interested in expanding the production of widgets. I will not attempt to detail this higher level problem but rather to go through the seven steps to decision and to note the reasons why each step becomes more difficult to take in the broader problem.

Step 1. Frame, in general terms, the context of the decision.

One very obvious difficulty in setting up the problem is that many more factors must be taken into consideration than was the case in the original example. If relocation of the plant is a possibility then there

will be at least a dozen important factors (transportation costs, labor supply, taxes, financing, etc.) which must be considered. Instead of a single objective there may well be multiple objectives. In general we must, therefore, deal with complex patterns of interrelated events.

Step 2. List the alternative courses of action.

One of the most difficult aspects of a higher level decision problem may be the preparation of a good list of alternatives. The range of possible actions is enormously broadened and it may be difficult to say whether a potentially important line of action has been omitted. The actions are much harder to compare because they may be quite different in character. In addition there is the difficulty that several variants of each general line of action may have to be considered. In the theory of Decision Making there are some general principles (such as admissibility) which may allow us to reduce the list of possibilities to a workable number by first eliminating subsidiary alternatives.

Step 3. Trace the possible outcomes of each course of action.

This step is likely to be a major undertaking for high level decision problems. The event chains that must be considered no longer have a simple form such as $Y_1\ Y'_2\ Y'_3$. To make matters still worse it will generally be necessary to consider not only immediate consequences but also long-term results of a given course of action. In general long-range prediction is a much tougher job than short-term forecasting.

Step 4. Determine the probability of each possible outcome.

Prediction in the widget production example is relatively simple because we are dealing with more or less repeatable events. In other words it is plausible to suppose that after recalibration we are essentially "turning the clock back and starting over again" so that we can easily amass a large body of relevant experience. Large scale actions, however, have relatively few precedents (i.e. similar experience) that can be used to determine probabilities. In other words, rather than dealing with repeatable events we tend to be dealing with unique events. To work with probabilities in such a situation it is necessary to develop considerably more sophisticated techniques. We may still try to construct mathematical models but these models must now involve the interrelationships between various relevant factors. In other words, the models required are at least equal in complexity to the ones employed in the more advanced work in chemistry and physics. Specially trained technicians may be needed to handle the job.

Step 5. Assess the desirability of each possible outcome.

Not only is it harder to obtain a numerical measurement of the desirability of the various outcomes but also the simple dollar and cents scale may sometimes be inadequate for the problem. Most major decisions are likely to affect the lives of a number of human beings and the success of the undertaking may be dependent on attitudes and value scales of these people. Although consumer preference studies and various methods of preference and taste-testing are being developed, we have a great deal to learn in these areas. In some ways it is the lack of adequate measurements of desirability that presents the principal barrier to the use of modern methods of decision making.

Step 6. Calculate the expected consequences of each action.

At present there is considerable discussion and even controversy concerning the theory for taking this step. Various devices for combining desirabilities and probabilities have been suggested. For example, if the probabilities are very rough estimates we may be able to devise criteria for assessing the expected consequences which are less sensitive to errors in the probability than the method of expected costs. Similar devices can be worked out if the desirabilities are not measured very accurately. However, these difficulties are largely technical and are likely to be surmounted by current research.

Step 7. Compare the expected consequences and select the most favorable course of action.

Here again there may be some essentially technical difficulties but these are not major stumbling blocks.

Step 8. Check up on the results.

This step is especially vital in the case of major decisions. Quite often a major decision can be broken down into a sequence of smaller decisions and if so, an early warning of the defects of some of the preliminary decisions may sometimes enable us to change our ways before it is too late. The concept of sequential decision is an important component of the newer methods of decision making.

I hope that I have not discouraged you by this somewhat pessimistic account of the difficulties encountered in making major decisions. I have mentioned the stumbling blocks so as to emphasize that decision making is not merely a matter of dumping in a bunch of figures and grinding out 100% pure decisions. The problem is intrinsically difficult and we have a great deal to learn. However, I do believe that as we attain proficiency in some of the simpler applications of Decision Making, we will be able to go on to bigger and better things. There is also the consolation that even where we cannot, as yet, carry through the seven steps to decision with accurate and reliable numbers at each stage it will still be true that the general principles will provide at least a rough guide, an approach to the problem of making the difficult decisions demanded by our highly technological culture.

Appendix:

Here are a few more details concerning the way in which Table I is constructed. If we recalibrate whenever a widget must be reworked the calculation of the exact probability that the chain ends with the nth widget is a rather tedious matter. The first widget gives us no trouble because the probability can be found directly from an ordinary Normal Integral Table (i.e. twice the area in the tail of the normal curve that extends beyond three standard units). Because the probability of a rework on the first widget is so small we make a negligible error if we use the normal table for the second widget as well (the only difference is that we look up the area for $3/\sqrt{2}$ standard units).

After the second widget the probabilities become harder to calculate. If Y_i is the measurement on the i-th widget (i=0, 1, 2, ...n) and we arbitrarily set $Y_0=0$, the formula for the probability of the chain ending on the nth widget can be written as:

$$P(Y_i') = \int_R \frac{e^{-\frac{1}{2}\sum_0^{n-1}(Y_i-Y_{i-1})^2}}{(2\pi)^{n/2}} \prod_1^n dY_i$$

where the multiple integral is taken over the region, R, where:

$$-3 \leq Y_i \leq +3 \qquad i=1, 2, \ldots, n-1$$

and

$$Y_n < -3 \text{ or } Y_n > +3.$$

These probabilities can be approximated by numerical integration. If we consider the probability that Y_n lies outside the -3 to +3 limits, <u>if</u> the previous Y's all fell inside the limits (i.e. the <u>conditional</u> probabilities) it turns out that the numerical values change rather slowly as n increases. Thus the <u>conditional</u> probabilities for the third, fourth, and fifth widget are respectively 0.079, 0.086, and 0.089. This fact enables us to get a reasonably good approximation from the simple formula given in the text.

Since some readers may wish to repeat the calculations in Table 1 for themselves, the entries in Column 3 have been calculated by using probabilities for widgets number one and two as explained above and thereafter using the factor 0.090 in the formula. Actually, of course, it would be better to use the results of numerical integration given above for the third, fourth, and fifth widgets. However, the effect is to overestimate the expected cost so the resulting appraisal is "conservative" in this sense.

REFERENCES

The only text on decision making which introduces the ideas without much mathematics is:

(1) Bross, Irwin D. J., <u>Design for Decision.</u> (New York: The Macmillan Company, 1953) 276 pages.

The best treatment of the topic (but one involving some very advanced mathematics) is:

(2) Blackwell, David and M. A. Girschick, <u>Theory of Games and Statistical Decisions.</u> (New York: John Wiley & Sons, Inc., 1954) 355 pages.

ACCEPTANCE SAMPLING - A DECISION MAKING TOOL

Prof. Gayle W. McElrath
Department of Mechanical Engineering
University of Minnesota

PREFACE

In practice we recognize that the chief function of an acceptance plan is to establish a criterion for accepting or rejecting lots. If the quality is all the same, the chances are that the plan will accept some of the lots and reject others. In this situation the accepted lots will be no better than the rejected. On the other hand, if the lots differ in quality, the plan will accept those lots of good quality more often than it will accept those of bad quality. In this situation, the average quality of the accepted lots will tend to be raised. If the rejected lots are given 100% inspection and then accepted, the average quality of the accepted lots is improved to the degree that the defective items are discarded.

It should be emphasized that a sampling plan alone cannot guarantee that the quality of the accepted product will be high. The quality of the accepted product also depends upon the quality of the submitted product and the disposition of the rejected lots, as well as the sampling plan. Acceptance sampling plans aim to give lot quality assurance; however, a very useful by-product can be the estimation of the average lot quality.

During the following discussion, we will be concerned only with those acceptance sampling plans which are used to accept or reject products submitted in "lots." Each item will be classified as either defective or non-defective. Lot quality will be measured by the percent of defective items in the lot. Such an acceptance sampling procedure is commonly called "acceptance by attributes."

SOME BASIC ESSENTIALS OF ACCEPTANCE SAMPLING THEORY

Since the modern theory of acceptance sampling involves the laws of probability, one of the basic essentials is that the sample must be a RANDOM SAMPLE. Most of us have been exposed to the term random sample until we have almost become immune to the expression. However, it is important to form a sample which arises from a process that "assigns to every member of the lot the same chance of belonging to the sample." It is very true that the inspector often faces definite physical difficulty in forming his sample; even so, he should strive to invent some procedure of selecting the sample without bias from all parts of the lot. Perhaps our greatest difficulty in making the sample a random sample is human laziness. We should not forget that the statistical theory fundamental to scientific acceptance sampling is based on the assumption that the sample be a random sample.

However, the fact that we have obtained a random sample does not guarantee an effective acceptance sampling program. There are other necessary essentials. There must be INSPECTION INSTRUCTIONS which define clearly the inspection tests and which minimize as far as possible any needless subjective judgment on the part of the inspector. In addition, an effective sampling program demands the quality of supervision that will insure that the items are carefully and impartially inspected, and

that the results are accurately recorded on the inspection record sheet. Failure to conscientiously meet these essentials will equally weaken the effectiveness of any acceptance sampling program.

Basic to the statistical theory of acceptance sampling plans is the variation of sample fractions defective when sample after sample is drawn from the same lot quality. All of us have observed that the sample quality is not necessarily that of the lot quality. In fact, if we would carry out an experiment of repetitive sampling from the same lot, we would discover that the frequency of occurrence of the sample fractions defective would take on a pattern. This pattern is called the SAMPLING DISTRIBUTION of the fractions defective for the given sample size.

An example will illustrate the point that we are making. Consider a lot of, say, 500 parts which is 10% defective. We will draw a sample of size 10 from this lot. We are interested in the probabilities of obtaining the different sample fractions defective. How often should we think of the sample as being better than, as good as, or worse than the lot? Consider the following table of values:

PROBABILITY DISTRIBUTION OF THE SAMPLE FRACTIONS DEFECTIVE
FOR SAMPLE SIZE 10, LOT SIZE 500, LOT QUALITY 10% DEFECTIVE

Number of Good Items in Sample	Number of Defectives in Sample	Sample Percent Defective	Probability of Obtaining Sample (Percent)
10	0	0	34.86
9	1	10	38.78
8	2	20	19.37
7	3	30	5.74
6	4	40	1.12
5	5	50	0.14
4	6	60	0.01
3	7	70	0.00
2	8	80	0.00
1	9	90	0.00
0	10	100	0.00

From the above table, we observe that only 38.78 percent of the samples will define the lot quality exactly, that 34.86 percent of the samples will be better than the lot quality, and that 26.38 percent of the samples will be worse than the lot quality.

We cannot hope to present the theory of probability and of sampling distributions. This information can be found in many of the standard quality control texts. However, we will mention that there are three ways of calculating probabilities for attributes inspection:

THREE WAYS OF CALCULATING PROBABILITIES FOR ATTRIBUTES INSPECTION

COMBINATORIAL probability distribution is used when calculating CHANGING probabilities. This occurs when the lot size is small enough so that the probability of drawing a defective changes substantially during the drawing of the sample size n. The lot size N is the predominant factor but the ratio of sample size to lot size n/N is also important.

The combinatorial probability formula which gives the probability of x or less defectives is given by

$$\sum_{i=0}^{x} \frac{C_i^{pN} \; C_{n-i}^{(1-p)N}}{C_n^N}$$

where N, lot size
n, sample size
x, number of defectives in the sample
p, fraction defective of the lot

BINOMIAL probability distribution is used when calculating CONSTANT probabilities. This occurs when sampling from a large lot size where the partial exhaustion of the lot by the sample does not significantly change the existent probabilities. When sampling from a conveyor or machine the assumption of indefinitely large lot size is usually correct.

The binomial probability formula which gives the probability of x or less defectives is given by

$$\sum_{i=0}^{x} C_i^n \; p^i \; (1-p)^{n-i}$$

where n, sample size
p, fraction defective of the process
x, number of defectives in the sample

POISSON probability distribution generally gives highly accurate APPROXIMATIONS to the binomial calculations when used for sampling inspection methods. The majority of sampling inspection problems and tables are based on the Poisson approximation because of the extreme simplicity involved in its calculation.

The Poisson probability formula which gives the probability of x or less defectives is given by

$$\sum_{i=0}^{x} e^{-np} \frac{(np)^i}{i!}$$

where n, sample size
p, fraction defective of the process
x, number of defectives in the sample

In summary, when the samples drawn are random samples, the theoretical laws of probability specify that chance alone will inevitably operate to give rise to "two types of wrong decisions". Some of the time, the sample information will indicate that we accept substandard lot quality, and then another part of the time, the sample information

will indicate that we reject acceptable lot quality. Any form of sampling will yield wrong decisions part of the time; however, the virtue of modern acceptance sampling techniques is that the risks of making these wrong decisions can be pre-assigned. The modern acceptance sampling plan can be designed to give the specified protection at desired economic levels. The more protection that is required of an acceptance sampling plan, the higher is the cost in terms of excessive sampling. The moral of the story is that there must be a desired balance between risks and costs.

SOME BASIC QUESTIONS TO ACCEPTANCE SAMPLING THEORY

What Types of Acceptance Sampling Plans are Available?

A SINGLE SAMPLING PLAN is completely specified by three numbers. The size of the lot N, the size of the sample n to be drawn from the lot, and number of defective units c that cannot be exceeded in the sample without rejecting the lot. However, if the Poisson formula is used to calculate the probabilities, the lot size need not be specified, only the sample size n and the acceptance number c are necessary. For example, if we have the following single sampling plan:

Type of Sampling	Sample Size	Acceptance Number	Rejection Number
Single	75	2	3

A random sample of 75 items is selected. If 2 or less defective items are found in the sample, the lot is accepted; if 3 or more defective items are found in the sample, the lot is rejected.

A DOUBLE SAMPLING PLAN can be illustrated as follows: The inspector takes a first sample of 50 items from the lot and inspects each of them. If he finds 1 or less defectives in the first sample, he accepts the lot. If he finds more than 4 defectives in the first sample, he rejects the lot. However, if the inspector finds 2 or 3 defectives in the first sample, he proceeds to take a second sample of size 100, and inspects them. Now if the combined 150 items contain less than 4 defective items, the lot is accepted, and if the combined number of items contains 4 or more defective items, the lot is rejected.

Type of Sampling	Sample Number	Individual Sample Size	Combined Samples		
			Size	Acceptance Number	Rejection Number
Double	1 2	50 100	50 150	1 3	4 4

A MULTIPLE SAMPLING PLAN may be described in the same manner as the double sampling plan except that the number of successive samples required to reach a decision of acceptance or rejection may be more than two.

Type of Sampling	Sample Number	Individual Sample Size	Combined Samples		
			Size	Acceptance Number	Rejection Number
MULTIPLE	1	20	20	+	2
	2	20	40	0	3
	3	20	60	1	3
	4	20	80	2	4
	5	20	100	2	4
	6	20	120	2	4
	7	20	140	3	4

+ Acceptance not permitted at this sample size

UNIT SEQUENTIAL SAMPLING PLANS permit items to be drawn and inspected one at a time. After each item is inspected a decision is made on the basis of the cumulated inspection results. The decision being either to accept or reject the lot or to continue sampling by taking another item.

<u>What Protection Does the Plan Afford?</u>

From the theoretical point of view, we know that lots with no defectives will always be accepted and lots 100 percent defective will always be rejected. Although we seldom know the exact fraction defective p of the lot submitted for acceptance sampling, we would still desire to have a picture of the plan's ability to discriminate between lots of different quality. This picture can be obtained and is called the OPERATING CHARACTERISTIC curve (OC curve) for the acceptance sampling plan. That is, if we assume the lot quality fraction defective to be p, then the OC curve will tell us the probability P_a of accepting this lot for our given sampling plan. The values of P_a are calculated by one of the probability formulas (usually the Poisson formula) mentioned in an earlier part of the paper.

Our discussion concerning the properties of OC curves will be limited to OC curves for single sampling plans. However, comparable remarks can easily be extended to include the OC curves for the other types of plans - the purpose of the OC curve does not change with the type of sampling plan.

The OC curve for the single sampling plan, $n = 75$, $c = 2$, is shown in Figure 1. The horizontal axis is the percent defective (100p) of the lot that is being submitted for inspection. The vertical axis is the probability P_a that the lot will be accepted by the sample. The respective values of P_a were found from Table G from reference 1.

PERCENTAGE SAMPLING is still a common type of inspection used to determine whether to accept or reject submitted lots. There are still those who think that the protection given by a sampling plan is constant if the ratio of sample size to lot size is constant. The OC curves uncover this mistaken idea of constant protection for constant percent sampling. For example, from Figure 2 we observe that a 3% defective lot will be accepted 73% of the time if we use a 10% sample from a lot of size 100, 53% of the time by using a 10% sample from a lot of size 200, 20% of the time by using a 10% sample from a lot of size 500, and only 4% of the time by using a 10% sample from a lot of size 1000. (The above probabilities were calculated by the combinatorial formulas.) This is a weakness in percentage sampling. That is, either too many or too few items are inspected for the desired protection - unless the lot just happens to be the correct size.

Fig. 1a. - Operating Characteristic Curve for the Acceptance Sampling Plan n = 75, c = 2

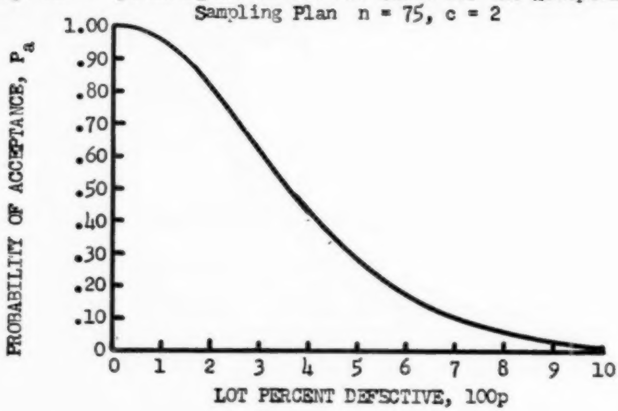


Fig. 1b. - Average Outgoing Quality Curve for the Acceptance Sampling Plan n = 75, c = 2

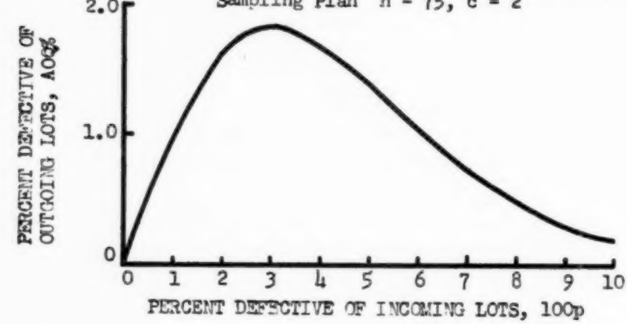


Fig. 2a. - Comparison of OC Curves for 10% Sampling Plans for Inspection Lots of 100, 200, 500 and 1000 Items

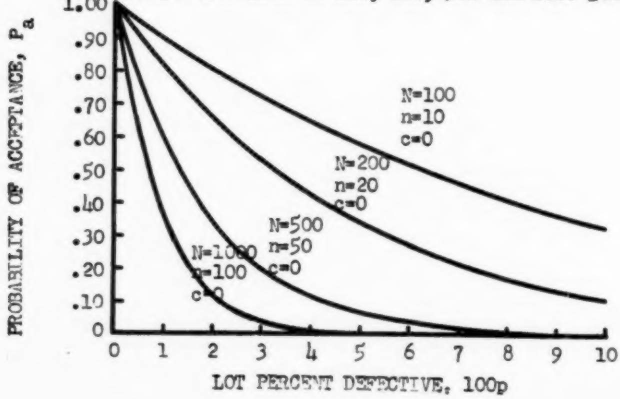
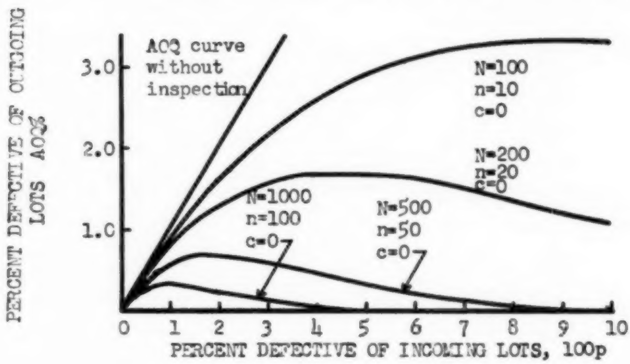
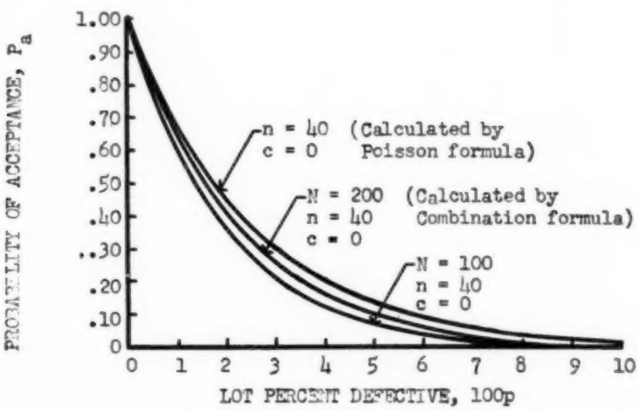


Fig. 2b. - Comparison of AOQ Curves for 10% Sampling Plans for Inspection Lots of 100, 200, 500 and 1000 Items



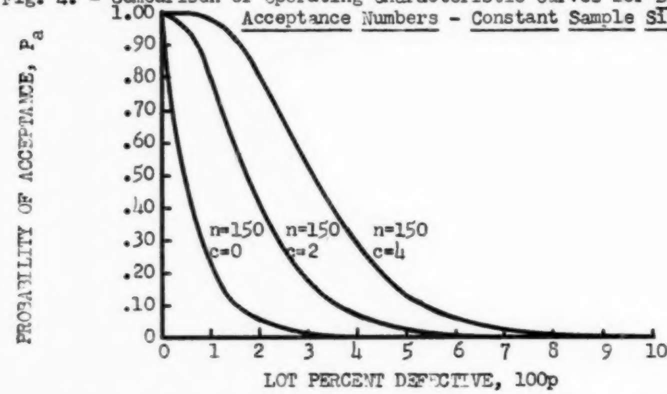
Many people in industry think that it is almost heresy to say that the ABSOLUTE SIZE of the sample is much more important to constant quality protection than the relative size of the sample compared to the lot size. For comparative OC curves which present a convincing argument of close agreement for fixed sample sizes drawn from different lot sizes see Figure 3.

Fig. 3. - Comparison of Operating Characteristic Curves for Fixed Sample of 40, and Acceptance Number of 0



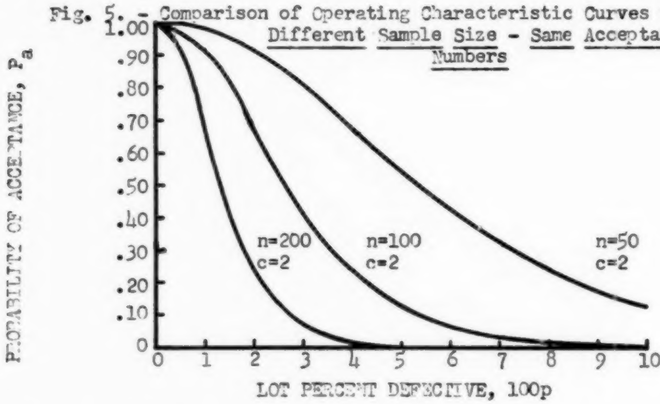
Let us investigate comparative OC curves when they are based on the SAME SAMPLE SIZE n but on DIFFERENT ACCEPTANCE NUMBERS c. It seems reasonable that as the acceptance number is decreased, the OC curve is lowered. That is, regardless of the lot percent defective, we are allowing fewer defectives in the sample to accept the lot. The effect is to tighten the plan. As c is increased, the OC curve raises. The effect is to make the plan more lax. See figure 4.

Fig. 4. - Comparison of Operating Characteristic Curves for Different Acceptance Numbers - Constant Sample Size



We may also be interested in the comparison of OC curves when they are based on the SAME ACCEPTANCE NUMBER but have DIFFERENT SAMPLE SIZES. Here again, we might foresee that as the sample size is increased, the OC curve is lowered. That is, regardless of the lot proportion defective p, there is increasingly a smaller chance of obtaining c or less defectives in the sample as the sample size increases. The effect is to tighten the plan. Similarly, as the sample size is decreased, the OC curve raises. The effect is to make the plan more lax. See Figure 5.

Fig. 5. - Comparison of Operating Characteristic Curves for Different Sample Size - Same Acceptance Numbers



The ideal sampling plan would discriminate perfectly between lots containing less than or equal to, say p_o fraction defective from lots containing more than p_o fraction defective. Such a plan would have an OC curve similar to that in Figure 6. Unfortunately, there is only one method of "sampling" inspection that will give this ideal OC curve, that is perfect 100 percent inspection. However, by JOINTLY VARYING THE SAMPLE SIZE AND THE ACCEPTANCE NUMBER, we can design a sampling plan which will discriminate between lot qualities as precisely as we please. See Figure 7. But immediately one might correctly remark that the greater precision is more expensive. And again we are forced to say that there must be an economic balance worked out.

Fig. 6. - Operating Characteristic Curve of a Plan Discriminating Perfectly at a Given Lot Percent Defective

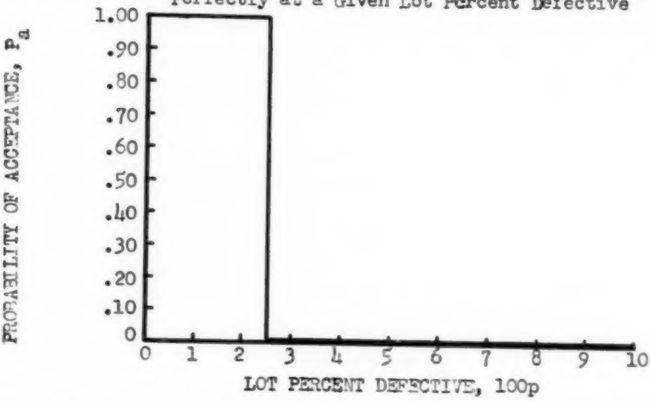


Fig. 7a. - Comparison of Operating Characteristic Curves for <u>Different Sample Sizes and Different Acceptance Numbers</u>

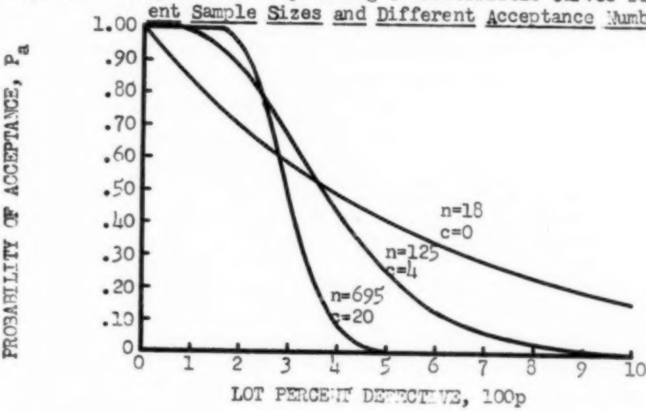
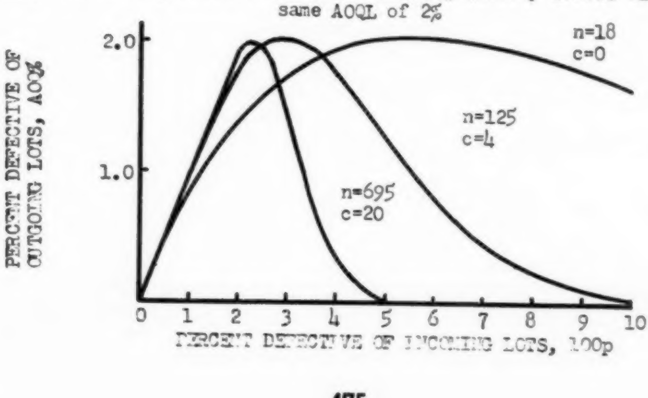


Fig. 7b. - Comparison of Average Outgoing Quality Curves with same AOQL of 2%



What is the Quality of the Product Passed into Stock?

We have stated before that the sampling plan alone cannot guarantee that the quality of the accepted product will be high. In fact, accepted lots passed into stock are almost the same quality that they were before inspection. That statement applied equally well whether the quality of the submitted lots was good or bad.

An acceptance sampling inspection procedure can indicate that the rejected lots be 100 percent inspected, and that the defective items be replaced by good ones. After this sorting operation these detailed lots will be passed into stock together with the accepted lots. This type of plan gives a definite assurance that the AVERAGE OUTGOING QUALITY (AOQ) of a large number of lots will not be poorer than a limiting percent defective called the AVERAGE OUTGOING QUALITY LIMIT (AOQL). This is one method of setting an upper limit on the percentage of defectives in the product that is finally passed into stock. The AOQ theory is essentially a process of dilution.

Consider a very simple derivation of a formula for the AOQ. Let all defective items found in the sample or in the rejected lots be made non-defective by either replacing, repairing, or by reworking. We will define:

p, lot fraction defective

Pa, probability of accepting the lot (P_a may be read from the OC curve)

N, lot size submitted for inspection

n, sample size taken from the lot

On the average, there are $p(N - n)$ defective items left in the accepted lots which will not be further inspected. And since the probability of accepting a lot is P_a, there will be, on the average, $P_a p(N - n)$ defective items passed into stock. Given the above conditions, we will define the AOQ as

$$AOQ = \frac{P_a p(N-n)}{N} \times 100$$

Figure 8 presents an illustrative diagram showing the "Theory of AOQ".

If the defectives found both in the sample and in the rejected lots are thrown out but <u>not</u> replaced, the formula becomes

$$AOQ = \frac{P_a p(N-n)}{N-np-(1-P_a)(N-n)p} \times 100$$

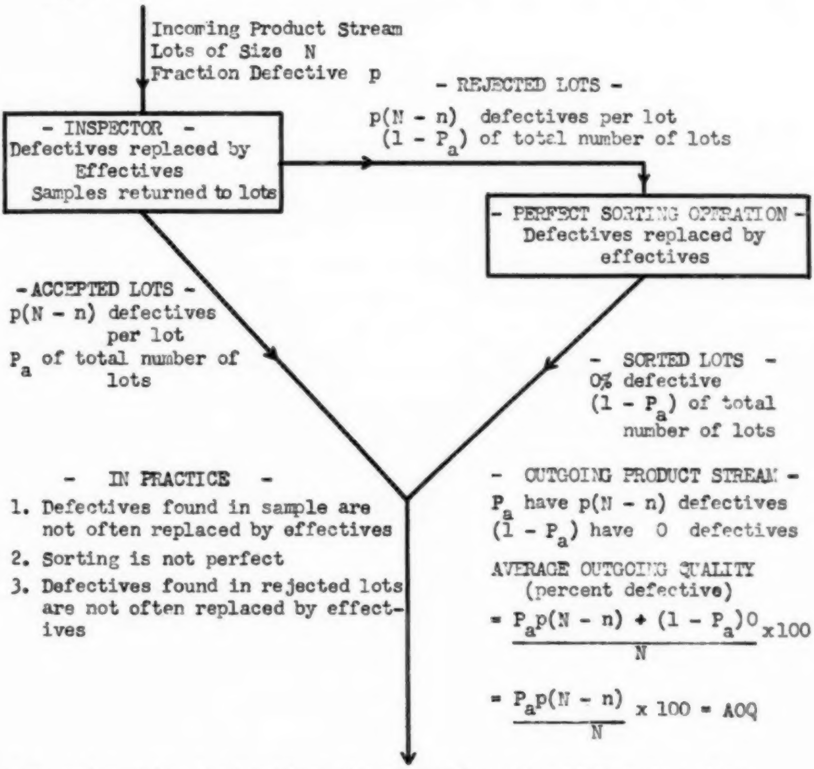
We reason that, on the average, np defectives will be thrown out of the samples and $(1 - P_a)(N - n)p$ defectives will be thrown out of the rejected lots. This means that there remains, on the average, $N - np - (1 - P_a)(N - n)p$ items per lot that finally go into stock.

The expression $AOQ = 100P_a p$ is a close approximation to the above formula when the lot size N is large and the sample size is relatively small. Whether the approximate formula can be tolerated can quickly be judged since the error of the approximation can be calculated easily.

We shall use the formula $AOQ = 100\ P_a p$ for determining our AOQ curves. These curves are plotted with the AOQ (percent defective) on the vertical axis, and the lot percent defective (100p) on the horizontal axis. The respective AOQL can be easily estimated by observing the maximum height of the AOQ curve. We observe that the AOQL value is a constant for a particular sampling plan and does not depend upon the incoming quality of the materials. It is of interest to notice that most of the lots will be accepted when the percent defective of the incoming material is low. As a result, the AOQ will be <u>low</u> in percent defective. Most of the lots will be rejected when the percent defective of the incoming material is high and therefore will be 100% inspected. As a result, the AOQ again will be <u>low</u> in percent defective. However, when the percent defective of the incoming material is between 0% and 100%, some of the lots will be accepted and others rejected; the expected AOQ for a given percent defective of incoming lots can be read from the graphs shown in Figures 1b, 2b, 7b, which are illustrations of AOQ curves.

THEORY OF AOQ

Fig. 8



For most practical situations the expression $AOQ = 100 P_a p$ is a close approximation to the above formula - especially when the lot size N is large and the sample size n is relatively small.

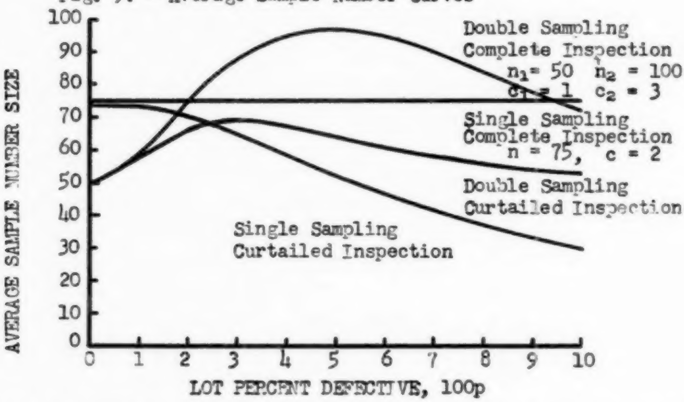
How Much Does the Plan Cost?

We cannot hope for a complete answer to this question because each company has its own set of cost factors which make each situation unique. Such items as cost of installing and supervising the plan, the cost of obtaining and inspecting the sample, and perhaps the cost of inspecting 100% should all be considered. In addition, to these very practical questions is the question of: how many pieces, on the average, is the plan going to require inspected? Of course, if the plan is a single sampling plan, and if the buyer would desire to estimate the supplier's quality level by a p-chart, then all of the sample should be inspected regardless of the number of defectives found. In this case for single sampling plans, the amount of sampling to be done for each lot is fixed at the sample size n itself.

However, if inspection of the single sampling plan is to be curtailed as soon as a decision of acceptance or rejection is reached, it is possible to derive a formula for the AVERAGE SAMPLE NUMBER (ASN) for a given percent defective p. That is, on the average, it will take a certain number of observations to make a decision on a submitted lot quality of fraction defective p. This procedure is called CURTAILED SINGLE SAMPLING.

In order to get a measure of the supplier's quality level when the plan is either double or multiple, a p-chart is kept only on the information from the first sample inspected. Inspection of later samples may be curtailed as soon as a decision can be reached. Unfortunately the theory necessary to construct the ASN curves is too detailed for our discussion but may be found in reference 2. The ASN curves for the single and double sampling plans previously considered are given in Figure 9 for comparative purposes.

Fig. 9. - Average Sample Number Curves



In a single sampling plan, if AOQ procedure is a part of the acceptance sampling, we may construct TAI curves which give a measure of the TOTAL AVERAGE INSPECTION per lot.

p, fraction defective of incoming lots

P_a, probability of acceptance

N, lot size submitted for inspection

n, sample size taken from the lot

then the expression for the TAI for a single sampling plan is given by

$$TAI = n + (1 - P_a)(N - n)$$

The TAI curve is plotted with the total average inspection on the vertical axis and the percent defective of incoming lots on the horizontal axis.

We can read from the above formula that a value of $P_a = 1$ gives a value of TAI = n. This sounds reasonable because material with no defectives will not be rejected, and the amount of inspection per lot will be just the sample size n. Again a value of $P_a = 0$ gives a value of TAI = N. To explain this we observe that every lot with all items defective will be rejected and therefore 100% inspected. In this case, the amount of inspection per lot will be the lot size N. Now if the percent defective of the incoming lots is between 0 and 100, the TAI curves give a measure of the total average inspection per lot and can be used to determine which sampling plan, among those offered, will give minimum inspection for different lot qualities. It must be emphasized that ingredient in the value for the TAI is the assumption that the rejected lots are detailed. Different TAI curves are given in Figure 11b.

FURTHER CLASSIFICATION OF SAMPLING PLANS ON THE BASIS OF TYPE OF PROTECTION

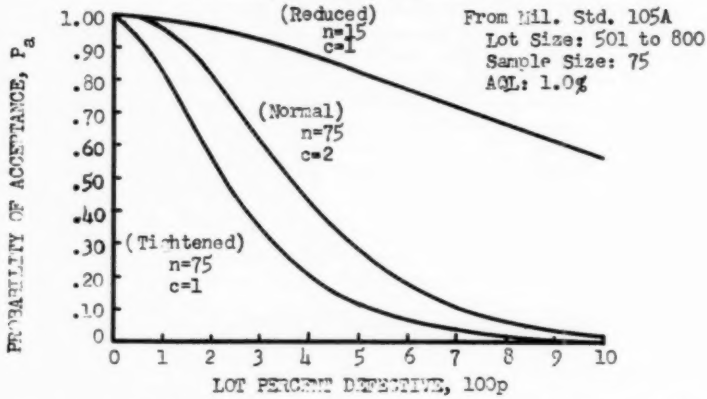
Suppose that we wish our plan to give specific protection against rejecting good lots. We shall think of good quality as that grade of material which is considered acceptable; therefore, the term ACCEPTABLE QUALITY LEVEL (AQL). Of course, it would be desirable to accept every lot quality better than or equal to the AQL. But this is the ideal; in sampling inspection we must run a risk of rejecting AQL lot quality. The OC curve for the sampling plan gives us the probability that AQL lot quality will be falsely rejected. This is the risk that the "producer" is taking that his submitted AQL material will be rejected. The probability of rejecting AQL material is called the PRODUCER'S RISK which is usually denoted by the Greek letter, α (alpha).

Suppose that we think of the AQL to be 1%. We find that our single sampling plan $n = 75$, $c = 2$ (Figure 1a) will falsely reject AQL lot quality with probability of 0.04. Technically, the probability of accepting AQL lot quality can be arbitrarily set; however, 90% and 95% acceptance of AQL quality are popular values. The sampling plans given in the MILITARY STANDARD 105A are classified by their AQL's.

When a producing unit or vendor has improved a process from an unsatisfactory level to a level substantially better than that considered satisfactory, it is fairly reasonable to expect a reward in the form of REDUCED INSPECTION. AQL inspection will do this without losing control of the process. One might observe that actually no inspection is necessary when the quality of the product is AQL or better. However, after a process has reached a very good quality level, inspection must be careful to exercise enough control over the process to maintain that quality level. If inspection is completely dispensed with, as could economically be done, the tendency for the process is to gradually deteriorate. A continuous check on the quality level given by the AQL inspection, and an informed production department tend to stabilize this improved process.

If the producing unit or vendor fails to keep the quality of the product sufficiently high, TIGHTENED INSPECTION can be installed in order to give added protection against accepting relatively low-quality product. Comparison of OC curves for reduced, tightened and normal inspection is given in Figure 10.

Fig. 10. - Comparison of Operating Characteristic Curves for Tightened, Normal and Reduced Inspection



Another feature that can be built into an acceptance sampling plan is that bad quality lots have a low probability of being accepted. We shall think of bad quality as that grade of material which we would like to reject. That is, there is some percent defective value such that we would like to reject all lots as bad as, or worse than, this given value. This value is known as LOT TOLERANCE PERCENT DEFECTIVE (LTPD). The probability of making the wrong decision of accepting lot quality of LTPD value is easily read from the OC curve. This probability is known as the CONSUMER'S RISK and is symbolized by the Greek letter β (beta).

Suppose we choose the LTPD value to be 7%. From our single sampling plan (Figure 1a), the probability that we will make the wrong decision of accepting lot quality of LTPD is 0.11. Again, the probability of accepting lot quality of LTPD can be arbitrarily set; however, 5% and 10% acceptance for LTPD are popular values. It should be remarked that every point on the OC curve has its own producer's and consumer's risk, but most often these risks are specified only for the AQL and LTPD.

LTPD inspection considers the problem of designing inspection plans where the primary purpose is the elimination of lots of highly unsatisfactory quality from the outgoing stream. The task of picking the proper LTPD usually is not difficult because the use of LTPD type inspection and the LTPD itself is normally dictated by an economical, engineering or psychological consideration that lots worse than a certain percent defective cannot be tolerated. As to the choice of the probability of rejection, that choice depends solely upon the nature of the defects and how much inspection can be economically paid for. Figure 11a shows a set of sampling plans with the same LTPD.

The two points (AQL, α) and (LTPD, β) uniquely determine a sampling plan. Suppose that the consumer does not want to accept lots which are 7% defective or worse any more than 10% of the time, and that he would

like the producer to send him at least 1% defective material or better. However, the producer wants a plan that will accept his 1% defective material at least 95% of the time. Now we have a sampling plan specified by the two points on the OC curve; namely, (1%, .95) and (7%, .10). We might label this inspection, AQL - LTPD inspection.

Fig. 11a. - Comparison of Operating Characteristic Curves for Single Sampling that have Approximately the Same LTPD of 5%

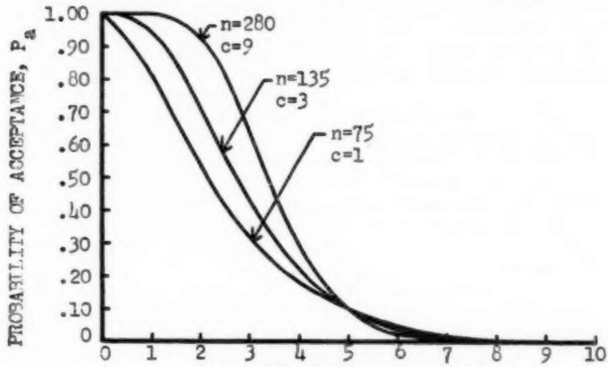
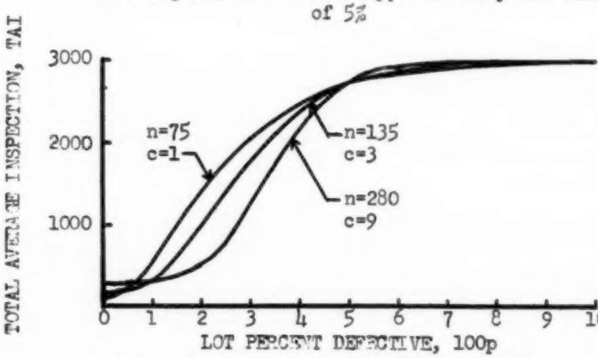


Fig. 11b. - Comparison of Total Average Inspection Curves for Single Sampling Plans that have Approximately the same LTPD of 5%



We have already considered the meaning of AOQL. Perhaps AOQL inspection is one of the most popular types of sampling inspection. It should be used in those places where the primary purpose is to control the average quality of the product leaving inspection. Nothing is said about the quality of any lot which happens to be either accepted or rejected. In fact, it is entirely possible to have a percentage of the lots passed into stock which are worse in quality than the specified AOQL. The important point to remember is that the AOQL feature aims to keep the average outgoing quality from exceeding a pre-assigned limit. Figure 7 presents different plans with the same AOQL.

Now consider very briefly what we might label as AQL - AOQL inspection and LTPD - AOQL inspection. The possibilities of these two types of inspection can be far-reaching. AQL - AOQL inspection guarantees the AOQL and also makes quite sure that lots as good as or better than the AQL are accepted. The features are particularly useful for the vendor who wishes to guarantee his customers an average quality limit and who wishes to protect himself by passing practically all the good lots produced. AQL - AOQL plans can be found in reference 2. Just as AQL - AOQL inspection is excellent for the vendor, LTPD - AOQL inspection can be ideal for the customer, if he wishes to hold an upper limit on the average quality that he sends into his plant and at the same time be reasonably certain that he will not pass poor lot quality.

Still another classification of sampling plans is that of AOQL - Minimum Total Inspection for a given fraction defective p. In this case the consumer or manufacturer, whichever it may be, has a plan which assures them that the material used or supplied will not, on the average, be worse than the designated AOQL; and at the same time will enjoy the benefit of low inspection costs under the usual operating conditions. Tables of Sampling plans which emphasize the AOQL - Minimum Total inspection are presented in reference 4.

The last combination to be considered is LTPD - Minimum Total Inspection for a given fraction defective p. This inspection procedure assures the user that he can be reasonably sure that very poor lots will not be passed into assembly; and at the same time he will enjoy the benefit of low inspection costs under usual specified conditions. Again we are fortunate to have tables of sampling plans which emphasize the LTPD - Minimum Total Inspection in reference 4.

SUMMARY

There exist different protection features which may be emphasized and built into an acceptance sampling plan. Although it is not possible to specify and get all of the features possible from the sampling plan, it is quite often possible to make an excellent compromise among the desired features and obtain a plan which will be close enough for all practical purposes.

The properties which have been discussed with respect to single sampling plans can be built into double, multiple, and sequential acceptance sampling plans. Comparative advantages and disadvantages in the amount of protection, administration, supervision, etc. should be taken into consideration when choosing an acceptance sampling program.

Correct mathematical theory is not enough to make a successful acceptance sampling program. There must be additional investments such as trying to take a random sample, carefully defining the defects, conscientiously inspecting the pieces, and accurately recording the results.

In the author's opinion, usually it is not necessary to be able to design the sampling plan, but it is necessary to become acquainted with the properties of the sampling plans in order to know their advantages and limitations. Most often satisfactory sampling plans can be selected from the many tables which are now available.

REFERENCES

1. Grant, Eugene L., "Statistical Quality Control", (2nd Ed.), McGraw-Hill Co., Inc., New York, 1952.

2. Statistical Research Group, Columbia University, "Sampling Inspection: Principles, Procedures and Tables for Single, Double and Sequential Sampling in Acceptance Inspection and Quality Control", McGraw-Hill Co., Inc., 1948.

3. Duncan, A. J., "Quality Control and Industrial Statistics", Richard D. Irwin, Inc., Chicago, 1952.

4. Dodge, H. F. and Romig, H. G., "Sampling Inspection Tables: Single and Double Sampling", John Wiley & Sons, Inc., New York, 1944.

5. "Military Standard 105A: Sampling Procedures and Tables for Inspection by Attributes", United States Government Printing Office, Washington, D. C., 1950.

6. Bowker, A. H., and Goode, H. P., "Sampling Inspection by Variables", McGraw-Hill Co., Inc., New York, 1952.

7. Molina, E. C., "Poisson's Exponential Binomial Limit", D. Van Nostrand Co., Inc., New York, 1949.

8. Romig, H. G., "50-100 Binomial Tables", John Wiley & Sons, Inc., New York, 1953.

MULTIPLE COMPARISONS WITH A STANDARD

Charles W. Dunnett
American Cyanamid Company
Research Division, Lederle Laboratories

I. Introduction

A problem which arises frequently in many fields is the comparison of treatment categories with a standard or a control. As examples of such a situation, consider the following: (a) An agriculturalist has several new varieties of corn which he wishes to compare with a standard variety in the hope that one or more will prove to be superior to the standard variety. (b) A pharmacologist has several drugs which have shown promise in the treatment of some disease, and he wishes to test them for the presence of some undesirable "side effect", such as a tendency to cause a rise in blood pressure. To do this, each drug is administered to a group of subjects, and the results compared with those obtained in a control group of untreated subjects. (c) An engineer is investigating the effect of changes in the manufacturing process of a radio tube in the hope of prolonging the average life of the tube. Several tubes are manufactured according to each process, and the results of life tests on them compared with similar observations made on tubes manufactured according to the standard process.

2. Comparison of a single mean with a standard.

As a concrete example, consider the following problem which is taken from Villars (1). Suppose we are interested in knowing whether treatment with a certain chemical results in a stronger cloth than that obtained by a standard manufacturing process. Three samples of cloth are obtained from the standard process to compare with three samples which have been chemically treated. The following table shows the pounds pull at which the specimens broke.

Breaking Strength (lbs.)

	Standard	Chemical
	55	55
	47	64
	48	64
Means	50	61
Variances	19	27

The chemical treatment appears to have had the desired effect, but it would be wise to apply a statistical test in view of the relatively large variability. The conventional procedure for comparing two mean values involves Student's t-distribution. We first compute the standard deviation, which is the square root of the average variance, $s = \sqrt{(19 + 27)/2} = \sqrt{23} = 4.80$. The standard error of the difference between the means is $s\sqrt{2/N} =$

$4.80\sqrt{2/3} = 3.92$. Finally, the "allowance" for the observed difference between the means is $ts\sqrt{2/N}$, where t is a factor obtained from tables of Student's t-distribution with $2(N-1) = 4$ degrees of freedom corresponding to the desired level of confidence. If 95% confidence is satisfactory, the required value of t can be read from the k = 2 column of table 1. For 4 degrees of freedom, it is t = 2.78, so that the allowance becomes (2.78) (3.92) = 10.9 in this example.

We can thus conclude that use of the chemical treatment will result in an increase in the breaking strength of $61-50 \pm 10.9$ = between 0.1 and 21.9 lbs.

When we make this statement, we are making a prediction regarding how much the breaking strength may be expected to increase on the average if we incorporate the proposed chemical treatment in the process. We have concluded that the increase is between 0.1 and 21.9 lbs., but this conclusion may be wrong. We can control the probability of it being wrong by the choice of the factor t - the way we chose it, we have a probability of 5% that it is wrong, i.e. that the increase is either less than 0.1 lb. or more than 21.9 lbs. To put it another way, if we apply this procedure to a large number of such situations, and make statements each time that the true difference between two means that we observed is between -- and --, 5% of these statements will be wrong. We can thus say that we are working at an error rate of 5% wrong statements. Of course, if we want a smaller error rate, say 2% or 1%, we can achieve it by going to the 2% or 1% columns in a table of Student's t.

In all this, we have assumed that the observations we have made can be considered to be representative of future observations which could conceivably be made. We cannot expect, for example, that our prediction will apply to any type of cloth if our observations were obtained with cotton. Furthermore, if the three specimens representing the standard procedure were cut from one bolt of cloth, while the three representing the chemical treatment came from another, it is possible that the observed difference in breaking strength reflects a difference between the two bolts rather than an effect of the chemical treatment. Use of the term "allowance" is intended to convey that only the sources of variation that have been designed into the experiment are allowed for. If there are other important sources of variability not allowed for, then the error rate may be considerably more than its nominal value.

3. Comparison of several means with a standard.

Now let us consider what happens when we have more than one treatment to be compared with the standard. Consider the following data, where the results of two further chemical treatments have been added to the previous data.

Breaking Strength (lbs.)

	Standard	Chemical A	Chemical B	Chemical C
	55	55	55	50
	47	64	49	44
	48	64	52	41
Means	50	61	52	45
Variances	19	27	9	21

Most statistics textbooks advise using the analysis of variance when there are more than two means to compare. We will try this technique, and then show that it cannot answer the questions the experimenter is probably most interested in asking.

We first compute the average variance, which is $(19 + 27 + 9 + 21)/4 = 19$. Then, the variance of the means is calculated and multiplied by N=3 to put it on a "per observation" basis; this gives 134. We can then set up the following analysis of variance table:

Source of Variation	Degrees of Freedom	Mean Square	F-ratio
Processes	3	134	7.1*
Replicates	8	19	

* Statistically significant at P=.05 probability level.

The ratio of the two variances is 7.1, which is beyond the 5% critical value of the F-ratio given in statistical tables. We can thus conclude that the processes are different. However, this may not be of much interest to the experimenter. He very likely knew beforehand that the chemical treatments would differ in their effect on the cloth. What he wants to know is "How much does each treatment affect the breaking strength of the cloth?". The analysis of variance per se cannot provide the answer to this question.

In the past, the experimenter has had to rely upon the so-called LSD (Least Significant Difference) procedure to answer this question. However, there is a pitfall which may ensnare the unwary investigator when he uses this concept. Basically, the LSD is simply an "allowance" pertaining to a difference between two means, calculated in much the same way as we did previously in comparing a single chemical with the standard. The standard deviation is the square root of the average variance, $s = \sqrt{19} = 4.36$. The standard error of a difference between means is $s\sqrt{2/N} = 4.36\sqrt{2/3} = 3.56$. Then, taking t = 2.31 from the k = 2 column of table 1 for 8 degrees of freedom, we get LSD = (2.31) (3.56) = 8.2. By the LSD procedure, this value is used as an allowance for the difference between any two observed mean values. If we used it in our example, we would conclude that chemical treatment A increased the breaking strength by $61-50 \pm 8.2 = 2.8$ to 19.2 lbs., and similarly for the other chemical treatments.

The objection to the use of the LSD procedure in comparing more than two means is its effect on the error rate. When we use the 5% value of t to compute the LSD, we will be working with an error rate of 5% wrong statements, which means that over a long period of time, 5% of our statements will not bracket the true value we are looking for. However, since we will make several such statements in one experiment, we will run a more than 5% chance of having a wrong statement in the experiment.

The following table indicates how the error rate on an experiment basis, i.e., the percentage of experiments in the long run which contain one or more wrong statements, increases with the number of treatments being compared with the standard. Of course, if the experimenter uses the LSD to make other comparisons in the experiment besides those with the standard, for example if he compares two chemical treatments, the error rate will be even higher.

k, number of means	Error rate
2	5.0%
3	8.8
4	11.8
5	14.4
6	16.6
7	18.6
8	20.3
9	21.9
10	23.3

In most cases, the experimenter must make some decision on the basis of his experimental results. He might be justified in feeling somewhat apprehensive if he had to make a decision on the basis of an experiment which had an appreciable probability of containing a wrong conclusion. The author believes that most experimenters would prefer to use a procedure which held the error rate on an experiment basis fixed at 5%, or some other suitable value. Table 1 gives a table of the factor t required to accomplish this in the case of comparing several treatment means with a standard. A more complete version of this table plus a corresponding table of 1% values and similar tables for one-sided comparisons have been computed by the author (2).

To illustrate the use of this table in our example, we compute as before the standard deviation, 4.36, and the standard error of a difference between means, 3.56. Then we take t=2.94 from the k=4 column for 8 degrees of freedom, and compute the allowance (2.94) (3.56) = 10.5. We are then entitled to make the following three statements, with the assurance that the probability is 95% that all of them will be simultaneously correct:

(i) The average breaking strength using chemical A exceeds that of the standard process by 61-50 $\pm$ 10.5 = between 0.5 and 21.5 lbs.

(ii) The average breaking strength using chemical B exceeds that of the standard process by 52-50 $\pm$ 10.5 = between -8.5 and 12.5 lbs.

(iii) The average breaking strength using chemical C exceeds that of the standard process by $45-50 \pm 10.5$ = between -15.5 and 5.5 lbs.

4. Comparison of error rate with that of LSD procedure.

It may be informative to illustrate numerically the difference in error rate between the LSD procedure and the procedure proposed here. Suppose that, over a period of time, 1000 experiments are performed with an average of 5 treatments being compared with a standard in each. Thus, in the 1000 experiments, 5000 comparisons are made.

If the LSD procedure is used, 5% or 250 of the 5000 comparisons will be wrong. However, according to the table given above, 16.6% or 166 of the 1000 experiments will contain wrong comparisons. On the other hand, if the procedure proposed in this paper is used, the experimenter will be guaranteed that only 5% or 50 of the 1000 experiments will contain wrong comparisons.

Incidently, if the 250 wrong comparisons were distributed at random among the 1000 experiments, there would be 226 experiments containing wrong comparisons instead of only 166. This illustrates that the wrong comparisons tend to occur in bunches rather than at random, as might be expected since comparisons made in the same experiment are positively correlated due to the presence of the standard mean in each.

5. Other multiple comparison procedures.

In many cases, the experimenter will want to compare the treatments with each other as well as with the standard. In that case, a multiple comparison procedure developed by Tukey (3) should be used. If the experimenter is also interested in making more complex comparisons, such as the average of a pair of treatments compared with the average of another pair, Tukey's procedure is also applicable but another multiple comparison procedure due to Scheffé (4) may be more efficient. Of course, Tukey's procedure or Scheffé's could also be used if the experimenter only wanted to compare each treatment in turn with the standard, but the "allowance" would then be larger than that obtained by using table 1 of this paper, so that some efficiency would be sacrificed.

6. Case where the standard value is known from past records.

In some cases, it will not be necessary to include observations on the standard in the experimental design, as information will be available from past records regarding the standard. For instance, if the standard represents a process which is in current use, it may be that so much data is available on the breaking strength of cloth produced by the standard process, that its average value is known very precisely. In this type of situation, we do not have to allow for any error in the mean for the standard, but only in the mean for the treatment compared with it. The allowance for a difference from the standard value is then $ts\sqrt{1/N}$, where t is taken from table 2 for k=4 means. This table was computed by Pillai and Ramachandran (5), where a table of one-sided values is

also given.

For illustrative purposes, suppose the value for the standard from past data is 50 lbs., and that we have an estimated standard deviation $s=4.36$ based on 8 degrees of freedom (using our previously obtained values). Then $s\sqrt{1/N} = 4.36\sqrt{1/3} = 2.52$, and since $t = 2.97$ by interpolation in table 2 for 4 means and 8 degrees of freedom, the allowance is $(2.97)(2.52) = 7.5$ (compared with 10.5 when we must allow for error in the standard value and use table 1).

References

1. D. S. Villars, "Statistical design and analysis of experiments for development research." Wm. C. Brown Co., Dubuque, Iowa (1951).

2. C. W. Dunnett, "A multiple comparison procedure for comparing several treatments with a control." Submitted to the Journal of the American Statistical Association.

3. J. W. Tukey, "The problem of multiple comparisons." Princeton University. Unpublished manuscript in private circulation.

4. H. Scheffé, "A method for judging all contrasts in the analysis of variance." Biometrika, 40 (1953):87.

5. K. C. S. Pillai and K. V. Ramachandran, "Distribution of a Studentized order statistic." Annals of Mathematical Statistics, 25 (1954): 565.

Table 1

5% values of t for comparisons with a standard
(standard value estimated from experiment)

d.f.	k, no. of means (standard included) 2	3	4	5	6	7	8	9	10
5	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
6	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
7	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3.86
8	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
9	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
10	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
12	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
14	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
16	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
18	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
20	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
24	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
30	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
40	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
60	2.00	2.27	2.43	2.55	2.63	2.70	2.76	2.81	2.85
120	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
inf.	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75

Table 2

5% values of t for comparisons with a standard
(standard value assumed to be known)

d.f.	k, no. of means (standard included) 2	3	4	5	6	7	8	9
5	2.57	3.09	3.40	3.62	3.78	3.92	4.04	4.14
10	2.23	2.61	2.83	2.98	3.10	3.19	3.28	3.35
15	2.13	2.47	2.67	2.81	2.91	2.99	3.06	3.12
20	2.09	2.41	2.59	2.72	2.82	2.90	2.97	3.02
24	2.06	2.38	2.56	2.68	2.78	2.84	2.91	2.96
30	2.04	2.35	2.52	2.64	2.73	2.80	2.86	2.91
40	2.02	2.32	2.49	2.60	2.69	2.76	2.82	2.86
60	2.00	2.29	2.46	2.56	2.65	2.72	2.77	2.82
120	1.98	2.26	2.43	2.53	2.61	2.68	2.73	2.77
inf.	1.96	2.23	2.39	2.49	2.57	2.64	2.69	2.73

"MACHINE TOOL CAPABILITIES"

Brent C. Jacob, Jr.
Chrysler Corporation

Every manufacturing business is a multiplant operation! You are fortunate if those plants where you make your scrap and do your rework are very small operations. Your other facilities may be making money, but your scrap and rework plants never do.

You have another profitless plant too! It is filled with unneeded equipment and useless operations.

Usual cost breakdowns speak of material, labor and burden. Scrap wastes all three. Rework wastes labor and burden. The plant filled with needless operations also wastes labor and burden. Insofar as its useless operations produce scrap, material is also wasted.

But you say you have no such plant as the "Division of Useless Operations". It is unlikely that you carry this division on your organization chart, anymore than you list "Scrap and Rework Divisions". Just the same, it is probably there.

If you keep normal books, you probably have a fairly good measure of your "Scrap Division" a poor or nearly non-existant measure of your "Rework Division", and no measure at all of your "Useless Operation Division".

All of this comes about much too naturally. In order to operate, you must account for the permanent shrinkage known as scrap. Any casualness about Operation Standards may permit untold amounts of rework to be buried in what appear to be 100% efficient operations. Measurement of machine tool capabilities occurs in relatively few plants; yet better utilization of those capabilities might permit reduction in the number of operations, or substitution of cheaper operations for costly ones.

Certainly, the factors in economical production of a quality product are numerous; many are difficult to measure. Yet without good measures of machine tool capabilities, how can an engineer set good specifications? How can the master mechanic tool the job when he doesn't know either the capability of the tool, or the extent to which that capability is going to be realized in use? How can management judge who is responsible for scrap, rework, excess inspection costs, etc. without such fundamental information? In all these cases the answers are negative. Machine tool capabilities are a vital requirement to economical design and manufacture of any product. Few people can honestly say that the economics of this concept has been pursued carefully, or even haphazzardly in their organization. Currently published

information leaves much to be desired in terms of economically sound approaches.

Review of records early in the application of SQC indicates many operations having process spreads five (5) times as great as machine tool capabilities. Not altogether rare are those cases where specifications which could not be met before application of SQC, were met consistantly with fewer and cheaper operations, simply by learning machine tool capabilities and providing needed controls to approach them.

Where machine tool capability spreads less than 50% of the specification spread, casual controls usually appear most economical. I repeat "appear", because I shudder to think of the cases where we overtooled at great expense, including those cases where we take two passes to do a job that could be done in one pass by less casual controls.

Most engineering specifications are the result of feedback from production, service, etc. Both the precision and the calibration of the feedback are subject to grave doubts. One case which illustrates this problem occurred in the production of our first V8's. Our cranks were superfinished at high cost. Oversize journals were lapped to size as a repair operation. We also scrapped many cranks for undersize journals. In studying the problem, we found the following interesting facts:

1. Cranks were being superfinished for excessive periods to bring journals down to size - almost on a custom basis.

2. To avoid undersize journals cranks were pulled from superfinishers before large journals were brought down to size.

3. Custom repair lapping was being used to bring the oversize journals to size without driving the others undersize.

4. Cranks were being ground too large generally and with great variation from one grinder to another.

5. Grinders were being instructed to approach an arbitrarily set lower limit of grinding size as closely as possible - but to make none under that size.

6. Finished crank journal sizes were distributed around the upper limit with 50% oversize journals before custom lapping (Figure 1).

7. After custom lapping, cranks were skewdly distributed, the tail being lapped off the high side so most cranks were from high limit $\neq$ .0002", with a long tail to well below low limit. There was high scrap from undersize journals. (Figure 2)

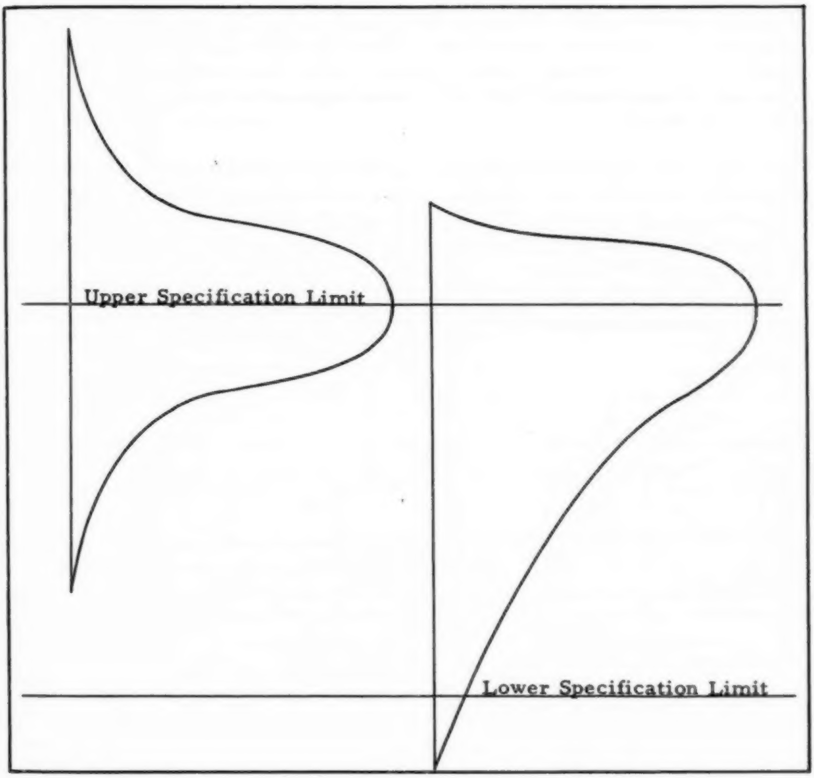


Fig. 1 Fig. 2

The problem "appeared" to be basically one of better control at the grinders, plus an educational program of high magnitude, not only for the grinders, but also for general crankshaft supervision to remove some of their "superstition".

When a competent grinder was told to grind "X" plus as little as possible and minus nothing, he aimed .0002" above X and seldom "got caught" below X because he only produced around 2 or 3% below X. When a poorer grinder received those same instructions he gave himself more leeway and aimed .0005" to .0008" above X and got by on the same basis. (Figure 3)

Some orderly technician had preceeded us, and determined that the superfinishing process removed from .0003" to .0007" on a normal cycle - more, and more varied on a longer one. Since the total specification was .001 finished size tolerance for use with standard bearing shells, grinders were instructed to grind no journals smaller than .0002" above high limit for finished journals. This was the "X" of Figure 3 to allow for superfinish. Since ground journals varied .0011" and superfinish removed up to .0007" production could plainly add and get .0018" which meant that the custom approach was "necessary" from their viewpoint.

Machine capabilities showed "as measured" values of .0007" realizable in grinding, and .0004" on superfinishing. Both were normally distributed and randomly associated so the square root of the sum of the squares, or a combined process variation of about .0008" would result.

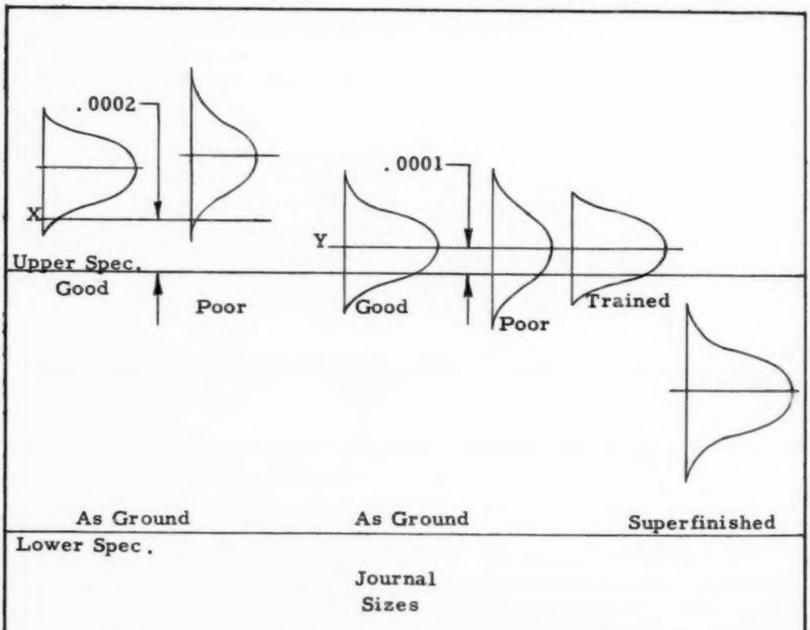


Fig. 3 Fig. 4

By a series of production tests, we proved that by instructing each grinder to aim at a "Y" of .0001" above high limit of finished cranks, and with a single normal pass through the superfinisher, the average size of journals was high limit - .0004". Variation was $\pm$ .0004". This resulted in a normally distribution range from high limit to .0008" below high limit or .0002" above low limit, which allowed enough for polishing for nick repairs, in the worst cases. (Figure 4)

At this time, we had just gotten our horse on 100% sawdust, as the old story goes, when we suddenly encountered a growing bank with mysterious knocks ahead of the repair hole. Our engineers were baffled because this knock, which showed some of the characteristics of a main bearing knock, appeared somewhat different.

I assured them that these were main bearing knocks, and proved my point. I cured several cases by inserting .001" undersize shells. Reinstallation of standard shells brought the knocks back.

The rest of the story goes like this. Main bearing knocks behave differently on V8's than on line engines, so our engineers, at this early stage in their own V8 experience, could be excused for their confusion. However, they admitted that in the face of our records, specified minimum clearances should be reduced .0002". Maximum clearance was then further reduced by changing the point at which .001" under shells would be selectively fitted. A small footnote states that they had always said the maximum clearances would knock, but had never so specified as to preclude knocks. They had let inspection reject knocks which we repaired by selective bearing fits. Here then, is the case of a feedback with a bias, or systematic error because inspection had consistently passed .0002" oversize journals about which engineering kew nothing. (Figure 2) The precision was also bad, because we had never told them our knock repair basis.

In this case when engineering at last received unbiased information, it responded by changing our specifications. By better specifications and better process controls our engines universally contained better bearing fits. As a sidelight, since only one division of the corporation had at that time so analyzed the problem, the Central Engineering Department dared not change the specifications to reflect the facts of life to the other divisions, as they would have increased their journal sizes instead of just accurately sizing them. This would have caused an epidemic of seized bearings. Chrysler Division obtained improved cranks cheaper by working honestly with the facts of life. All "Custom" operation had ceased, scrap was reduced, and repairs dropped sharply. Everyone benefited.

In another case a ream plus hone operation was replaced with a well controlled ream only. One expensive operation was eliminated. Size was held enough better that an operation which had previously been high in both scrap and rework went to negligible rework and 1/2% scrap, along with elimination of 100% inspection. This was a saving in our

scrap division, our rework division, and our useless operations division. We did others like this and so can you. It seems reasonable to expect a 500% return on investments in such effort. By improving the economic balance of these control methods we hope to do even better in the future.

At this point I rest my case for the desirability of knowing machine tool capability. I will, however, share with you some of the techniques we have tried, and will also try to evaluate the effectiveness of some of these approaches.

There are numerous ways of measuring machine tool capabilities. Most surveys state these capabilities "as measured". Relatively few isolate measurement errors to infer the actual process capabilities, devoid of measurement variations. It should be obvious that calipers and a vernier scale would give us a different opinion of lathe capability than would electro-limit gages.

With an "undisturbed" process, consecutive pieces vary dimensionally. Minute differences in hardness, stock removal, spindle deflection, movement of ways, temperature etc. are some of the random causes. Such random causes usually combine so that the variation in sizes is approximately normally distributed. (Figure 5) That means that about 2/3 of the pieces are contained in the central 1/3 of the total variation; about 95% are in the central 2/3 of the total spread; and those larger than the mean size are pretty well balanced by those smaller than the mean.

Now if we replace a tool, and do not "exactly" restore the tool to the same location, this whole variation will be moved larger (or smaller) as a result of the new tool location. Such changes as this are not always normally distributed. (Figure 6) In fact we might have just as many in the outer third as we have in the central third of tool locations. Or we might have 3/4 of them larger than the mean and only 1/4 smaller.

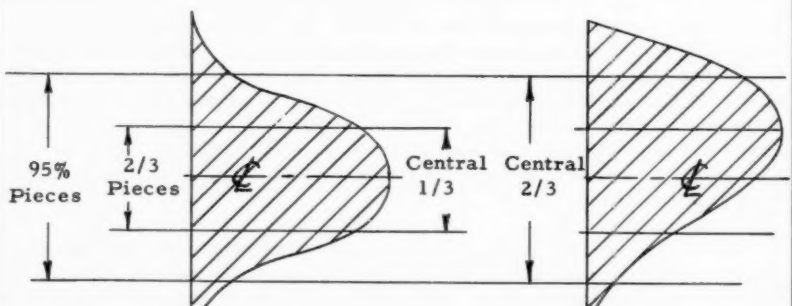


Fig. 5

Fig. 6

In any event the long range process spread is wider than the "undisturbed" process spread as a result of tool setting errors, changes, and other non-random disturbances to the process. (Figure 7)

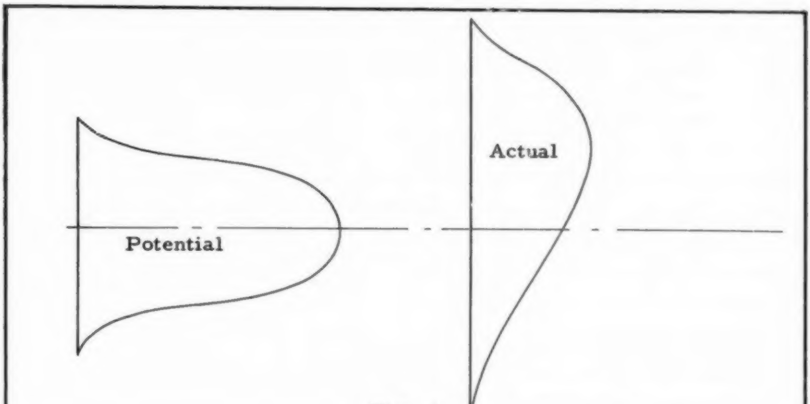


Fig. 7

The random causes of "undisturbed" process variation are usually very difficult to isolate and reduce. They normally require a basic process revision to change. The non-random disturbances are more likely to be separately recognizable and separately controllable to some degree. For instance, a better tool setting technique would help to make the long range process spread more like the "undisturbed" process spread.

Now if we could provide ourselves with enough facts, we could select that combination of basic processes and disturbance controls so that we could get the lowest total cost for meeting each specification.

Are you interested? Of course you are!

There are a number of ways to determine process performance. We are interested in both the undisturbed performance (Potential) and in the long run process performance including its disturbances, (Actual). We make or lose money on the actual. Improvement in actual, however, depends heavily on knowledge of the potential.

There are five (5) fairly common analyses used to survey processes:

1. A large sample of not less than 500 consecutive pieces may be plotted in production order. (Figure 8) This is expensive, easy to interpret in one sense, yet to a trained person less obvious than some cheaper methods.

2. Samples of not less than 50 pieces may be plotted as histograms. (Figure 9) This shows some things not obvious by the first method but loses the effect of order of production unless a coding technique is used.

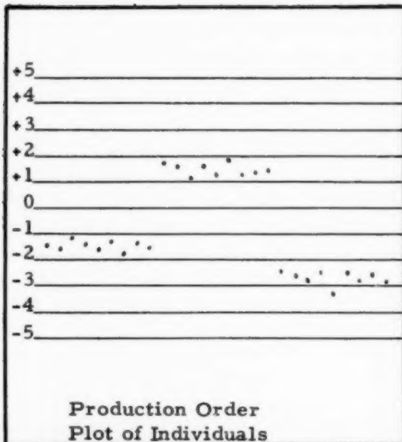


Fig. 8

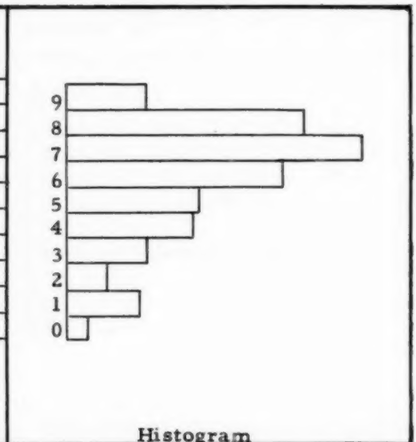


Fig. 9

3. Samples of less than 50 need the use of probability paper to establish distribution shape. Use of this paper permits surprisingly good interpretation from samples of 10 or even less, and increases our knowledge of large samples over that obtained by simple histograms. Time sensitivity is no different for a given sample size than histograms.

4. Use of average and range charts provides a great deal of information, both of a time sensitive nature and as to distribution width and centering. The shape of the distribution, however, is only assumed to be normal. Whereas this is usually an acceptable assumption it may occasionally conceal important facts.

5. The multivary chart helps to locate the relative significance of (a) within piece variation (b) short range variation and (c) long range variation. (Figure 10)

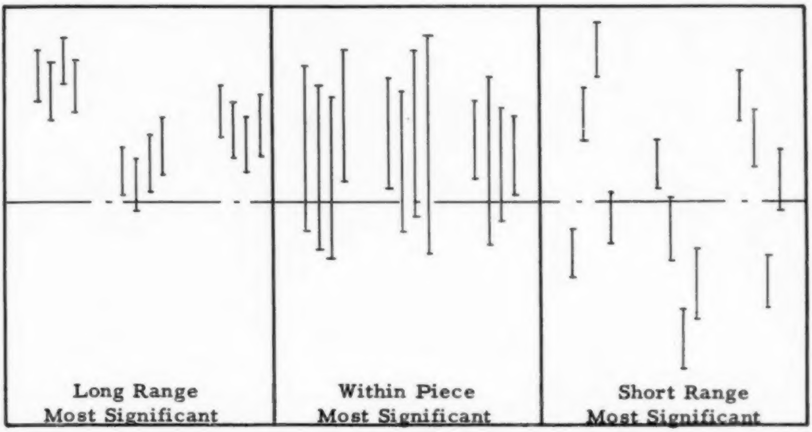


Fig. 10

Of these five (5) methods, only three (3) seem useful, and only two (2) seem normally applicable.

Multivary charts, by definition, force a look at variation within a single piece, such as taper, out-of-round, bellmouth etc. The balance of the analysis is weak.

It would appear better to routinely check within-piece variation without recording, except where this variation is obviously significant. Use of the average and range chart then permits observing a number of small samples which can appraise both the short and long range variation. If range observations are generally in control, $\overline{R}$ from samples consisting of consecutive or "undisturbed" operation may be used to compute a 6 sigma process potential, "as measured".

Use of probability paper permits simple analysis of measurement data for actual spread implied by the sample. The shape of the distribution is shown with increasing clarity as the sample size increases. (Figure 11)

Since potential or undisturbed variation is in many cases normally distributed, use of $\overline{R}$ for its determination has the merit of short runs. These reduce the likelihood of disturbances, average a large number of observations for added precision, and therefore tend to provide a fairly dependable picture of the potential.

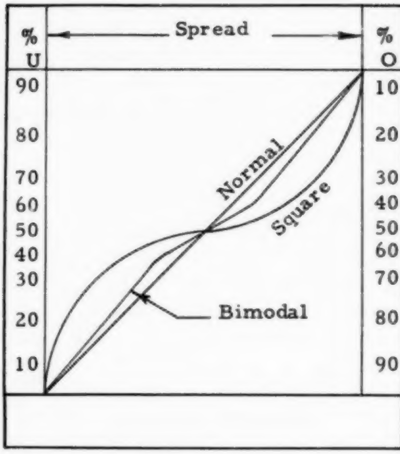


Fig. 11

Cells	Over #	Over Cum.	Over %	Under Cum.	Under %
+10					
+ 9	1	1	0.5		
+ 8	2	4	2.0		
+ 7	5	11	5.5		
+ 6	13	29	14.5		
+ 5	19	61	30.5		
+ 4	24	104	52.0		
+ 3	18	146	73.0		
+ 2	7	171	85.5		
+ 1	5	183		17	8.5
0	3	191		9	4.5
- 1	2	196		4	2.0
- 2	1	199		1	0.5
- 3	3	200			

Fig. 12

Because disturbances may well be other than normally distributed, use of the total data from the chart, replotted on probability paper gives a more reliable picture of the long run distribution, or actual.

Arthur Bender's arithmetic procedure for preparing data for probability paper analysis is so simple it should be used. (Figure 12) He collects the measurements in cells. He then accumulates the totals of all "above and left entries" to the right. A final total then equals twice the number of observations used. Thus 10 observations provides a total of 20; 25 a total of 50 etc. When this total is 100, each cell cumulative is directly in percent for use on probability plotting. If the total is 50, double each cell cumulative for percent; if total is 20 multiply each cell by 5 for percent etc. This makes sample sizes of 10, 25, 50 100 especially simple to use.

To know true process facts, the amount of "process spread" contributed by variability in measurement should be separated. This can be done by setting the gage to the master before each sample, then rechecking the gage against its master afterwards. The difference is recorded as R for the gage. $\bar{R}$ for the gage may be computed into 6 sigma spread for the gage when at master size. Ten-observation samples in the same location of the same piece for each of several sizes of pieces can be separately plotted on probability paper. If the spreads of the "eyeballed" lines, and the $\bar{R}$ derived 6 sigma spread, agree with their averaged spread within ± 25%, use the mean spread of all these as gage error spread. If these vary substantially more than ± 25% have the gage repaired or replaced, and start over on the whole investigation which involved that gage.

Assuming the gage spreads compared acceptably, true process spread may be inferred as equal to:

$$\sqrt{(\text{observed process spread})^2 - (\text{gage spread})^2}$$

Where the gage spread exceeds 25% of the observed process spread, the gage is poorly suited for process control or measurement.

Let us not forget those lovely dollars though. This technical labyrinth is useful. It is not an end in itself! - - So let us now see a moneymaking approach.

If we look at our scrap records and pick the parts having the largest dollar value of scrap in excess of 1% we already have one high return basis. High rework items yield another. Items which comprise a high service complaint basis may be another. Any items frequently processed and inspected against salvage limits yield yet others.

Apply average and range charts of 2 piece consecutive samples, randomly spread over a period of two weeks for a total 100 pieces inspected on each worthwhile case.

After 50 pieces have been charted, apply control limits on averages and ranges, with the control limit for average centered on specification mean rather than process mean.

As each subsequent sample is posted, immediately investigate for causes of any out-of-control, have corrections made, and record the causes.

When the 100 pieces have been inspected, analyze them on probability paper, by plotting the points and "eyeballing" the best fit straight line through points between the 10% and 90% points. Observe the degree of fit of these points, and also those outside the 10% and 90% points.

If the inside points fit well, the distribution is probably reasonably normal. If points outside the 10% - 90% range extend well beyond the ends of the range encompassed by the "eyeballed" straight line, they probably represent "wild shots" from poor operation.

If the spread between the .15 and 99.85% of the "eyeballed" line are less than 150% of the 6 sigma spread showed by the $\bar{R}$ analysis your control is above average and may be hard to improve. If it is 200% or over, it represents sloppy operation which should be easy to improve.

If the $\bar{R}$ calculated 6 sigma spread is over 2/3 of the specified tolerance spread, careful control is mandatory. If it is equal to, or greater than specified spread you need basic process improvement, or a revised specification if you hope to succeed. If it is substantially less than 50% of the specified tolerance spread you may be applying excessively expensive processing.

All the foregoing analyses are based on use of "as observed" spread, using gages whose spread is below 25% of the observed potential spread.

When industry becomes armed with machine tool capability facts, products can be more economically designed. Tools can be better specified knowing that machine tool vendors can provide us equipment more reliably matched to those specifications, and that production can be expected to realize a greater share of that capability. Managements can more certainly measure the effectiveness of their teams.

With the scrap, rework and useless operation divisions shrunk in size, the consumer can expect to share our management gains.

Once we know good economic specifications and process capabilities we can properly start to decide on whether we should spend more for tools and less for controls or vice-versa on each operation. This will let us achieve each specification at least cost.

FILL CONTROL IN THE CANNING INDUSTRY

C. B. Way
Green Giant Company

The need for the maintenance of proper fill of food products cannot be overemphasized. Slack filling results in poor acceptance of the product and overfilling is uneconomical and may be damaging to the product. Certainly a housewife who opens a can of food and finds it only half full will be dissatisfied. Likewise, she will be dissatisfied with a can of chicken soup for instance, which has no chicken in it. These are real problems in the food industry and have undoubtedly occurred in every plant where food is packaged. Many complaints of this type never reach the packer, but surely the housewife will be wary of the brand with which she was dissatisfied when she makes her next grocery purchases.

The poor economics of overfilling is clearly illustrated by the example of a baby food line running at a speed of 1000-5 oz. cans per minute. Overfilling by only one-fourth of one ounce would mean the loss of 3000 cans per hour. This is no small item and could mean the difference between profit and loss.

Processors who pack in narrow neck glass bottles such as catsup, syrups and liquors have a problem in that small variations in level of fill produce a very poor appearance on the retailers' shelves. This problem may be aggravated by non-uniformity of bottle volume.

In such products as cake and pie mixes, proper fill is essential to the end results since the package is used as a unit in the recipe. The uniformity of quality may be greatly affected by the fill in products such as beef stew, chicken noodle soup, etc., in which the individual ingredients are added separately to the can. There are also some products, particularly those cooked under agitating processes in which proper fill is essential to adequate sterilization. Over- or underfilling of these products can result in high spoilage losses.

Food products vary considerably in physical characteristics and thus present a number of fill problems. Several different types of fillers have been developed to handle these products. The principal types are the <u>headspace fillers</u> used for liquid products (juices, beverages, etc.) which fill to a predetermined headspace controlled by a displacement pad or by the location of the vent; the <u>plunger</u> or <u>piston fillers</u> used for liquid or semi-solid products (hash, cream style corn, pumpkin, etc.) which premeasure the product and discharge it to the container either forcibly or by gravity; the <u>volumetric fillers</u> used for small granular products (peas, diced vegetables, lima beans, etc.) which premeasure the product and discharge it to the container by gravity; the <u>hand packers</u> used for large granular products (whole vegetables, fruits, etc.) which use the container itself as the measuring unit; and the <u>scale type fillers</u> used for dry products (cereals, cake mix, flour, nuts, etc.) which weigh the material either in the containers themselves or in cups which discharge to the containers.

While the physical properties of the products and the types of fillers vary considerably the basic methods of evaluating the fillers and carrying out the statistical control of them are essentially the same for all types of products and fillers.

FILLER CHARACTERISTICS

One question to which the manufacturer and processor alike would like an answer is, "How precise will the filler fill?" or "When filling to a certain average weight or volume, how many cans (or what percent) will be more or less than this average by any given amount?" In other words, they are interested in knowing the distribution of fill-in weights which can be expected under a given set of conditions.

The manufacturer is interested in this information for obvious selling purposes. The food packer is interested from the standpoint of fill control. The packer is also, of course, interested in buying the most "precise" filler consistent with cost. It might be mentioned here that buyer and seller should be on common terms when talking about filler characteristics. The seller may state that his machine will consistently fill to 0.01 oz. meaning "on the average." The buyer could misinterpret this to mean that every container will be within ± 0.01 oz. of the average when actually some might vary as much as ± 0.5 oz. This example may be an exaggeration but does point out the need for being on common bases when discussing this subject.

The determination of this distribution of weights or "fill characteristics" is a simple matter and has been outlined in more detail in another article [10]. Briefly, it is done by drawing a number of consecutive containers (100 samples simplifies the calculations) from the line and taking the required weight, headspace or volume measurements. The standard deviation is then determined either graphically or by calculation. This should be done several times under the same conditions and under varying conditions of speed, level of fill and maturity of product (where it is a factor) in order to verify the results and to demonstrate the effect of variation of these factors.

Filler characteristics have been determined by several companies on a large variety of items and some of this data is presented in Table I. These figures are presented to give an indication of the characteristics of different fillers on a variety of products. They are meant to serve only as a basis to which others can compare their results and should not be used without verification. Some of the figures have been confirmed by more than one source. Others are the average of rather widely varying results.

Without going too much into detail there are some interesting points which might be brought out from this table. The absolute variation is nearly always higher in the larger can sizes, but it should be noted that the relative variation (coefficient of variation not shown) is lower. This is characteristic of most fillers and products. Where sufficient data is available, the range of observed standard deviations is given and is often quite wide. For instance, the standard deviation of the net weight of 303 peas varied from .06 - .42 oz. This is partly due to the fact that all the pea data has been grouped together for this purpose including several sizes and maturities of peas and several types of fillers run at different speeds.

There is considerable variation among the standard deviations of the product or drained weight of various products. These seem to be more or less proportional to the ease of filling. Note that whole beets, sliced beets and spinach which are normally packed on hand packers are somewhat higher in this respect. However, the variability of the net weights

TABLE I
FILLER CHARACTERISTICS

Product	Can Size	Standard Deviation: Net Weight	Product or Drained Weight	Sauce or Brine Weight
Peas	303	.14(.06-.42)*	.18(.09 - .45)	.18(.11 - .36)
	8 oz.	.13(.07-.23)	.12(.07 - .19)	.15(.08 - .46)
Kidney Beans	300	.14	.12	.13
Red Beans	# 2	.16	.10	.14
Hominy	# 2	.16	.12	.13
Pork and Beans	300	.12	.08	.11
Whole Kernel Corn	12 oz.	.30(.15-.60)	.17(.11 - .23)	.20(.10 - .47)
	#10	.77(.57-.96)	.67(.58 - .76)	1.05(.94 -1.15)
Beets, diced	303	.3	.6	.5
	#10	1.3	2.2	1.1
Lima Beans	303	.3	.9	.5
Peas and Carrots	303	.2	.4	.25
	#10	.5	1.5	1.5
Spaghetti	300	.2		
Beets, whole	303	.3	.5	.35
	#10	.8	1.8	1.5
Beets, sliced	303	.3	.7	.5
	#10	.8	3.0	2.0
Spinach	303	.5	.6	.4
Popcorn (Volumetric Filler)	10 oz.	.5		
Flour (Hand)	2,5,10#	.5		
Flour (Automatic)	2#	.25		
	5,10#	.33		
	25#	.7		
Cereals & Cake Mixes		.13		
Juices	6 oz.	.06		
	# 2	.12(.09-.15)		
	46 oz.	.30(.13-.39)		
Soups	8 oz.	.11		
	$10\frac{1}{2}$ oz.	.12		
Oil SAE 30	Quart	.06		
20	Quart	.10		
10	Quart	.14		
Lard	3#	.46		
Ketchup		.09		
Cream Style Corn	303	.15		
	8 oz.	.11		
Hash	300	.10		
Strained Baby Foods	6 oz.	.04		
Chop Suey	# 2	.40		
Chow Mein	303	.19		
Tomato Puree	8 oz.	.16		
Apple Sauce	5 oz.	.03		
Horsemeat	300	.15		
Dog Food	300	.16		

*Ranges given in parenthesis

is not particularly high which is probably due, in large measure, to the leveling affect of the brine addition.

Fillers, themselves, have not been shown to vary much in uniformity of fill on granular or non-homogenous products. The product, itself, seems to be the controlling factor. However, speed has a considerable affect on this uniformity, the tendency being to fill more irratically at very high speeds due to spillage and lack of time for the pockets to discharge completely. This has also been reported by packers of soups and juices where the high speeds tend to increase the spillage (and reduce the average fill). In some cases, improved design of liquid fillers has improved the uniformity of fill, however.

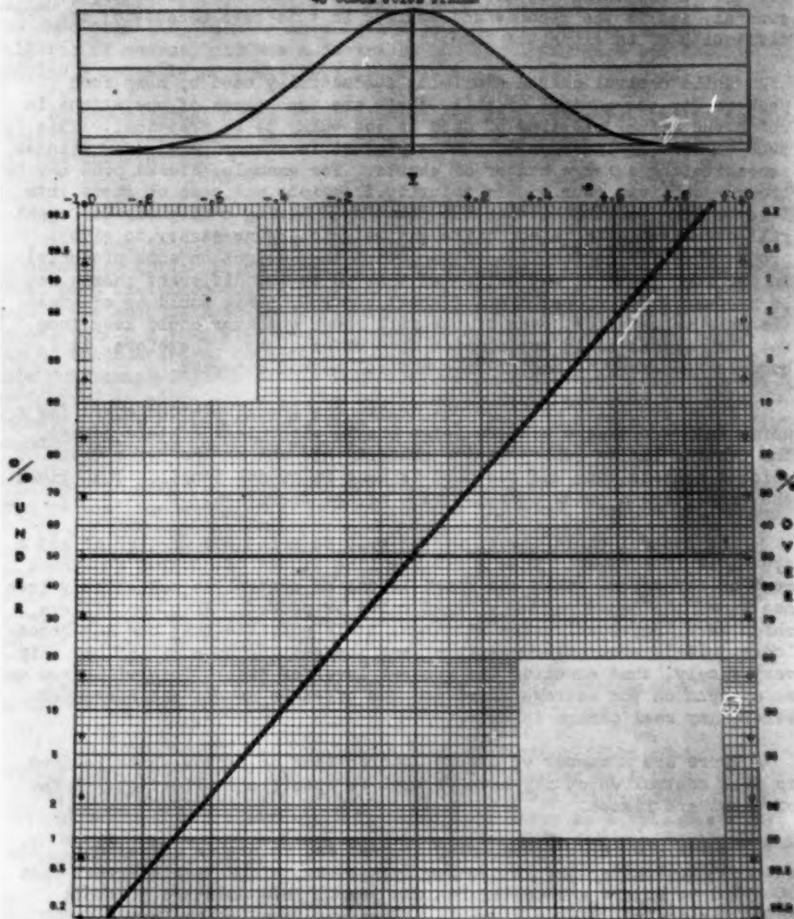
Having determined these characteristics it becomes important to present the data in such a manner that the management can quickly "see" how the filler fills. A method suggested by Bender [1] demonstrates a very effective way of doing this. His method involves the use of probability paper (which may have been used to determine the standard deviation) modified to include both the probability line and the distribution curve. An example is shown in Figure I. The distribution curve is given at the top and the probability curve below. From the probability curve the percentage of containers which can be expected to fall above or below any given weight can be read directly. For example, one may wish to know what percentage will fall below label weight (46 oz.) when the average is 46.48 oz. Reading down the chart at -.48 and across to the left hand margin it is seen that six percent will fall below label weight. Conversely, this chart can be used to determine at what level the filler should be set in order to fill any given percentage above or below a predetermined weight (or volume).

CONTROL

Once the characteristics of the filler are determined, the setting up of a control program is a relatively simple matter. Many packers are now using standard control charts ($\bar{X}$ and R charts) set up either from the known machine variability or from sample ranges. Three standard deviation (3-sigma) limits are most commonly used. The relevant methods are clearly outlined in many statistics texts and are not repeated here.

Some product fills are controlled by headspace measurements, some others by volume measurement, some by piece count, and many by weight measurements. Whatever the measurement, the procedure is about the same. A small sample (4 - 10 containers) is periodically taken from the line, measured, recorded, averaged and the range calculated. In order to save a calculation, the average is sometimes eliminated and the total used for control. In some cases it has been found unnecessary to consider the range, but this is a most useful control in many operations, particularly where one valve or pocket may become partially (or completely) plugged resulting in the slack filling of one container per revolution of the filler head. This might not throw the average out of control but evidence of faulty operation would be indicated in the range. This will be most effective if the sample includes all the valves or pockets. Fillers with ten or less valves can be covered in one sample; fillers with more can be covered in two or three samples. For example, on a 30 valve filler, ten consecutive containers starting with valve number 1 would be the first sample, the next ten would start with pocket number 11 etc. Thus, if the sampling period is twenty minutes, the entire filler would be covered every 40 minutes.

FIGURE I
FILLER CHARACTERISTICS
46 OUNCE JUICE FILLER



Sample sizes larger than ten or smaller than four are not recommended. Schrock (12) gives some very good reasons for this. Very small sample sizes necessitate too frequent calculations and large samples may be too infrequent to give good control. If the characteristics of the filler are markedly non-normal, the distribution of the averages of samples smaller than four will probably be non-normal and the 3-sigma limits will be invalid whereas for samples of four or more, the distribution of sample averages will be essentially normal and the 3-sigma limits will be valid. Small samples are also less sensitive to small shifts in process levels than larger samples. It is important to have some logical basis for selecting a sample size, and, as has already been suggested, for

fill control, it can be based on the number of valves or pockets in the filler. While sample sizes up to 15 are considered satisfactory, in general, ten is the recommended maximum in this case because of the difficulty of handling the samples.

While control charts are being successfully used by many food packers for the control of fill, there are some types of operations in which the actual plotting of data is too bulky to be practical. This is true of filling lines on which the product is changed every few minutes necessitating a large number of charts. For example, Alaska peas may be broken down into four to six maturity fractions and each of these into five or six fractions within the plant each having a slightly different fill characteristic, i.e., different weights are necessary to give proper eye-fill (which is more important than weight on some products). In one such plant it was estimated that 60 to 100 different charts would be necessary for proper $\overline{X}$ and R chart control. This would be of questionable value since, even if the fill check operator could keep them up, the foreman could not examine them effectively. Therefore, it becomes necessary to modify this control method somewhat.

A few companies which do not desire the actual plotted chart for a permanent record have mounted large charts with clear acetate covers. The control limits (which may be changed) are set on the cover with colored acetate tape and the plot is made in grease pencil. This gives the inspectors a quick over-all view of the operation.

The control chart itself serves only to point out whether or not a system is in control. When a filling line is out of control the cause must be corrected. When the range is out of control as has already been mentioned the cause may be a plugged valve or pocket. Also, on fillers which have individual valve settings, it is possible that one may become loose. It is also quite possible that the setting on a filler may slip very slowly, thus changing the general level of fill. This will show up as a trend on the average chart and can often be caught and corrected before any real damage is done.

There are a number of checks, corrections or improvements related to fill control which may have general or specific application. A few of these are listed.

Scales should be clean, easy to read and in proper adjustment. One source [2] indicates that in one New York county alone, scale errors cost $543,000 to buyers and sellers in one year. This amounted to about $6700 per scale. A West Virginia packer was found to be losing $1200 per day because of a weighing error. These are but two of many incidents of this type.

The use of properly installed and operated vibrating equipment is often helpfull in filling granular products. Transfers between fillers and closing machines should be checked. Rough transfers often spill product causing both waste and non-uniform fill. Frequent stopping and starting is also a common cause of poor fill.

A variety of electronic weighing devices are now available [7] and the use of X-ray headspace control has been developed [3,5,11]. These are used to eliminate the over- and underfills when it is desirable or necessary to obtain a narrower range than the filler can give. They are also used occasionally to indicate faulty operation of the filler.

Variable glass volume is often a problem. Some packers have contacted the glass manufacturers to see what can be done to improve this and some promising results have been obtained(10). Headspace juice fillers, of course, produce a desirably uniform appearance but on particularly valuable products this alone is insufficient.

Proper deaeration and tight connections to prevent reaeration are necessary in many products to eliminate entrained and/or dissolved air and prevent foaming.

IMPROVEMENTS RESULTING FROM STATISTICAL FILL CONTROL

In order to justify the installation of a fill control progam it is necessary to demonstrate definite improvements. A number of companies which have installed such a program have shown some very concrete savings. One such case is that of a stew line(10). The machine was found to be delivering a very wide range of weights (of a very expensive ingredient) which averaged considerably higher than the standard weight for this particular ingredient. A fill control program reduced the average by 5/16 of an ounce while at the same time greatly improving the uniformity. Over a single production run this saving was sufficient to fill an additional 950 dozen cans.

Another case was that of a dry package in which a fill control program reduced the average fill from 12.19 to 12.06 oz. The amount below 11 7/8 oz. was reduced from 7.6 to 5.5% and the amount above 12 1/4 oz. was reduced from 22.1 to 5.6%. On another line the average was reduced from 8.07 to 8.06 oz. which does not seem particularly great, but those under 7 7/8 were reduced from 9.8 to 3.8% and those over 8 1/4 were reduced from 12.6 to 5.4%. In each case the average was reduced slightly and the uniformity greatly improved.

Many other lines and plants have shown similar improvements as a result of the installation of a sound fill control program.

In conclusion, it has been shown that statistical fill control methods have invaded the food packing industry insofar as fill of container is involved and have been able to bring about improvements in fill uniformity and product savings. These methods serve as a valuable aid to operators and management but are <u>not a substitute</u> for properly trained and supervised personnel.

ACKNOWLEDGEMENTS

The auther wishes to express his sincere appreciation to the following for their co-operation in assembling this material: L. C. Berth, The Larsen Company; C. D. Trombold, Campbell Soup Company; I. MacPhail, H. J. Heinz Company; O. C. Voelker, Oconomowoc Canning Company; J. H. Houck, Libby, McNeill and Libby; L. A. Koritz, California Packing Corporation; L. Newsome, Continental Can Company; R. E. Burkholder and H. L. Link, Food Machinery and Chemical Corporation; and P. O. Lysne, General Mills, Incorporated.

REFERENCES

1. Bender, Arthur Jr., "A Graphical Method of Determining the Characteristics of a Frequency Distribution," Delco Remy Division of General Motors, Anderson, Indiana.

2. Canner, "Don't Let Weighing Errors Cut into Your Profit," Canner, 119 (26) 11-13 (1954).

3. Cram, G., "New Beams Monitor Fill Levels and Lift Barrier on Packaging Speeds," Food Engineering, 47 (June 1953).

4. FE Staff, "Air-Gap Weighing Cuts Overfill," Food Engineer, 62 (March 1953).

5. Goodman, David, "Checking Liquid Levels with X-Rays," The Brewer's Digest, 50 (May 1953).

6. Hegerarty, W. L., "Handling of Short Fills," The Brewer's Digest, 57 (January 1953).

7. Jenanyan, Edward H., "Save Overweight Losses, Prevent Underfills," Food Processing, 40-42 (January 1955)

8. Lee, E. P., "Getting Optinum Performance from Filling Equipment," Modern Packaging, 137 (June 1953).

9. Merrill, R. L. and Hecox, Wm., "Automatic Statistical Control Saves on Package Overfill," Food Engineering, 55 (May 1951).

10. National Canners Association, Proceedings of the Technical Sessions, January 23-27, 1954, Brittin, W.A., "Filling Semi-Solid Products," Edwards, H. Ivor, "Filling Semi-Solid Products," Link, H. L. and Dobson, H., "Fillers for Granular Products," MacPhail, Ian, "Filling Liquid Products," McKinley, E., "Filling Liquid Products," Way, C.B., "Fill Control of Granular Products."

11. Schreiber, H., "Fill Check by X-Ray," Modern Packaging, 197 (April 1953).

12. Schrock, E. M., Quality Control and Statistical Methods, (New York City, Reinhold Publishing Corporation 1950) Chapter VI.

Note: Not all these references have actually been used here. They are listed as being of possible interest.

MULTIPLE DECISION PROCEDURES FOR RANKING MEANS[1]

Robert E. Bechhofer
Cornell University

Introduction:

Most of the textbooks on statistics (including the best ones) attempt either implicitly or explicitly to classify all of the problems of statistics into one or the other of two categories: problems of statistical estimation, and problems of the statistical testing of hypotheses. Thus, for example, procedures are given for estimating (by point or interval) the unknown mean and/or variance of a normal distribution; and similarly, procedures are given for testing hypotheses concerning the unknown mean and/or variance of such a distribution. These procedures are well known and it is not our purpose to discuss them here.

Recently statisticians have realized that many of the important problems of statistics cannot be classified into either of the above categories. Appreciation of this fact has led to a reappraisal of the entire fabric of statistics--its formal framework, objectives, and philosophy. Out of this intensive self-scrutiny has developed the science of Statistical Decision Theory, a new mathematical theory which provides a rational basis for making decisions in the face of uncertainty. This new theory recognizes the fact that the purpose of data collection is to provide information which can be used for making decisions. Since the data obtained are variable, and hence of a statistical nature, there is a definite probability (however small) that any decision made using these data will be incorrect. Statistical Decision Theory weighs the possible economic losses associated with making incorrect decisions and then provides procedures which tell the experimenter which decision to make (from among the several possible decisions which are available in a given problem) based on the data at hand. Looked at in the above light, the problems of statistical estimation and the statistical testing of hypotheses are special cases of the more general problems that can be handled by this new theory.

The formalization of the concepts of Statistical Decision Theory is due to the late Abraham Wald who pioneered in the fundamental research in this area. The major fruits of his research are summarized in his book Statistical Decision Functions (7). However, this book was written for the mathematical research workers in the field and is highly theoretical and abstract. Since very few expository articles have appeared on the subject of decision theory, most of the users of statistics are unaware of the practical implications of this new approach to problems. This communication failure is particularly unfortunate because progress in the science of statistics depends upon a healthy interchange between practitioners who can define their problems in a meaningful way and theoreticians who can provide useful solutions to these problems.

Our objective in this present paper is to stimulate the interest of applied statisticians in the decision-theoretic approach to problems. We hope to do this by proposing a simple practical decision problem, and then indicating the considerations involved in finding a solution to this problem. We shall first pose the problem in general terms and then consider its precise statistical formulation. The complete statistical theory underlying our solution is presented in (1).

[1]Research supported in part by the Office of Naval Research and in part by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

Statement of the practical problem:

Suppose that it is desired to purchase a lot of steel bolts, and that several suppliers offer their product at essentially the same price. Suppose further that the characteristic of bolts which is of major importance is their tensile strength, a lot of bolts being considered more desirable the higher the average tensile strength of the bolts in that lot. Thus if it were known which lot contained the bolts with the highest average tensile strength, it would be known which lot to purchase.

In order to determine the lot averages exactly (neglecting the possibility of measurement error) it would be necessary to subject all of the bolts in all of the lots to a tensile strength test. However, it is obvious that (aside from cost considerations) this course of action must be ruled out immediately since the tensile strength test is a destructive test, and the complete information necessary to determine the lot averages exactly can be obtained only at the price of complete destruction of the lots. Clearly, in a situation such as this there is no alternative but to settle for incomplete information even though by doing so one runs the risk of making an incorrect decision (that is, of purchasing a lot which is not the best one). The only recourse is to take a random sample of bolts from each of the lots, and then using the results from the tensile strength tests on the bolts in each sample to make an inference about the tensile strength of the untested bolts in the remainder of the lot. Obviously this is a statistical problem because the results obtained from the tests on the bolts in the random sample from each lot will depend on which particular bolts from each lot were included in the sample. The objective then is to devise a statistical procedure which will tell the experimenter which decision to make (that is, which lot to purchase), the procedure having the property that the probability of making an incorrect decision will be controlled, in some sense, at a prescribed level.

Before proceeding it is important to emphasize that this is a decision problem, and not one of estimation or of testing hypotheses. That is, based on the available evidence it is desired to make one of the k decisions listed below:

Decision 1: Purchase the first lot
Decision 2: Purchase the second lot
⋮
Decision k: Purchase the k^{th} lot

In the language of Statistical Decision Theory such a problem is called a k-decision problem. The problems solved by the ordinary statistical tests of hypotheses which are presented in the literature are called 2-decision problems; this is so because one of two decisions (either to accept the hypothesis under test or to reject it) is made on the basis of the data. Any statistical procedure which is designed to handle these k-decision problems ($k \geq 2$) is called a Statistical Multiple Decision Procedure.

Statistical formulation of the problem:

It is assumed that the characteristic, tensile strength, is normally distributed, and that the tensile strength measurements are statistically independent--also, that a random sample of N bolts can be taken from each of the k lots ($k \geq 2$), the measurement for the j^{th} bolt in the i^{th} lot being denoted by X_{ij} ($i = 1,2,\ldots,k$; $j = 1,2,\ldots,N$). The true average tensile strength and the true variance of tensile strength for the bolts of the i^{th} lot are denoted by μ_i and σ^2, respectively

$(i=1,2,\ldots,k)$. It is further assumed that the μ_i are unknown, and that the common variance σ^2 is known. The ranked μ_i are denoted by $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]}$, and the difference between the i^{th} ranked average and the j^{th} ranked average by $\delta_{i,j} = \mu_{[i]} - \mu_{[j]}$ $(i,j=1,2,\ldots,k)$. Lastly, it is assumed that it is not known which lot is associated with $\mu_{[i]}$ $(i=1,2,\ldots,k)$, that is, it is not known which lot is "best," "second best," etc.

With reference to the above assumptions, the one concerning normality is not critical, and fairly large departures from normality can be tolerated with little effect on the results provided that the sample size N is not too small; however, lack of statistical independence affects the results in an unpredictable way. The assumption of a common known variance may not be warranted in particular problems unless a previous history with similar products indicates this to be the case. References to procedures which can handle certain problems in which this assumption is violated are given at the end of this paper in the section on Generalizations; however, it is important to point out that no procedure is available to handle problems in which absolutely no information is available concerning the variances.

The reader should not get the impression at this point that the procedures to be presented are applicable only to the tensile strength problem described above. On the contrary, once the statistical model has been provided, the reader is free, if he so desires, to interpret the above symbols in terms of any problem of his own choosing provided that the statistical model is also appropriate for that problem. For example, in an agricultural problem the X_{ij} can be interpreted as the yield/acre of the j^{th} plot sowed with the i^{th} variety of grain, the problem being to choose the best (i.e., highest yielding) variety; or in an ordnance problem the X_{ij} can be interpreted as the distance traveled by the j^{th} projectile in the i^{th} lot, the problem being to choose the best (i.e., longest range) lot.

General considerations underlying the procedure:

How would any statistically-inclined experimenter attempt to solve the problem of which lot is best? Most likely he would take a random sample of N bolts from each of the k lots, measure the tensile strength of the kN bolts, and for each lot compute the average tensile strength of the bolts in the sample. Then he would decide that the best lot was the one associated with the sample which had the highest average tensile strength.

What, if anything, is wrong with the above procedure? Nothing, except that the experimenter does not know what proportion of the time he will make a correct decision if he follows this same procedure time and time again in similar situations. He certainly feels that the larger the sample size N, the greater his proportion of correct decisions; also, the larger the difference $\delta_{k,k-1}$ between the unknown averages of the best and second best lots, the greater his proportion of correct decisions. However, he also knows that if the difference $\delta_{k,k-1}$ is very small, his proportion of correct decisions will be high only if N is very large.

Now it is clear that in many practical situations it may not be worth while to attempt to distinguish between the best lot and the second best lot if the difference $\delta_{k,k-1}$ is very small. This is so because the loss associated with making an incorrect decision may be small and/or the cost associated with guaranteeing a high proportion of correct decisions may be prohibitive. In fact, in many such situations it is possible to specify a constant $\delta^*_{k,k-1}$ which is the smallest

value of the difference $\delta_{k,k-1}$ that is worth detecting. That is, whenever $\delta_{k,k-1}$ is less than $\delta^*_{k,k-1}$, the experimenter is indifferent as to whether the best or the second-best lot is chosen; however, whenever $\delta_{k,k-1}$ is equal to or greater than $\delta^*_{k,k-1}$, the experimenter desires to choose the best lot a high proportion of the time. These general considerations suggest the properties that our procedure should possess, and we are now in a position to state the problem precisely.

Goal, specification, requirement, and procedure:

Goal: The experimenter's goal is to select the best lot, that is, the lot associated with the highest average $\mu_{[k]}$.

Specification: It is assumed that the experimenter can specify two constants before experimentation starts. These are:

1) The smallest value, say $\delta^*_{k,k-1} > 0$, of the difference $\delta_{k,k-1}$ that is worth detecting, and

2) The smallest acceptable value, say $P^* < 1$, of the probability P of achieving this goal when $\delta_{k,k-1} \geq \delta^*_{k,k-1}$. (It is clear that for the goal considered, the specified value P^* should be greater than $1/k$ since a probability of $1/k$ can be realized without taking any sample.)

Requirement: The experimenter requires that the procedure to be used guarantee that:

$$\text{Probability}\left[\text{Correct decision} \,\middle|\, \delta_{k,k-1} \geq \delta^*_{k,k-1}\right] \geq P^* ,$$

that is, that "the probability of a correct decision is to be equal to or greater than P^* whenever the true, but unknown, difference between the largest and second largest average is equal to or greater than $\delta^*_{k,k-1}$."

Procedure: The procedure which guarantees this requirement can be described as follows:

1) Enter Table I (at the end of this paper).

   a) The appropriate column is determined by k (the number of lots).

   b) The appropriate row is determined by the specified constant P^*. Representative values of P^* are given as row headings.

2) Set the entry obtained from Table I equal to $\sqrt{N}\,\lambda$ where $\lambda = \delta^*_{k,k-1}/\sigma$ and σ is the known population standard deviation. Solve this equation for N.

3) Take a random sample of N bolts from each of the k lots where N is the smallest integer equal to or greater than the solution obtained in 2), above. (If N turns out to be prohibitively large then the experimenter must specify a smaller P^* and/or a larger $\delta^*_{k,k-1}$.)

4) Measure the tensile strength of the kN bolts, and for each lot compute the average tensile strength of the bolts in the sample.

5) Select as the best lot the one associated with the sample which has the highest average tensile strength.

It is important to point out that the above procedure tells the experimenter how to determine the sample size in a rational way (for aside from this contribution the procedure is exactly the same as the one which any reasonable experimenter would follow); and this sample size is dictated by the constants, $\delta^*_{k,k-1}$ and P^*, which were specified by the experimenter. The emphasis has been placed on designing the experiment, and thus the experimenter is forced to state his requirements before experimentation starts; the "analysis" of the data obtained from the experiment simply consists of computing and ranking the k averages.

We shall now give an example to illustrate how the procedure is to be used.

Example:

Suppose that it is desired to decide which of five lots of bolts is best with respect to the characteristic average tensile strength, and that it is known from past experience with similar lots that the standard deviation of tensile strength for such bolts is approximately 1000 psi. How many bolts must be tested from each lot in order that the probability of a correct decision will be equal to or greater than 0.70 whenever the true difference between the largest and second largest average is equal to or greater than 500 psi?

In the above we have $k=5$, $P^*=0.70$, $\delta^*_{5,4}=500$ psi, and $\sigma=1000$ psi. Entering Table I in the column headed $k=5$ and in the row headed $P^*=0.70$ we obtain the entry 1.6614. Since $\lambda=\delta^*_{5,4}/\sigma=$ 500 psi/1000 psi $=0.5$, we obtain the equation $0.5\sqrt{N}=1.6614$. Solving this equation for N we obtain $N=11.04$ and conclude that 12 bolts must be tested from each lot.

Generalizations:

The procedure discussed in this paper is referred to as a single-sample multiple decision procedure because the final decision as to which lot is best is made on the basis of a single sample from each lot. The theory underlying the single-sample procedure is given in (1); this reference which also considers more general goals (such as selecting the best two lots or selecting the best three lots) contains a table (from which our Table I was abstracted) necessary for the application of the procedure to problems involving some of these more general goals. The single-sample procedure assumes a common known variance of tensile strength from lot to lot. If the variances are unequal from lot to lot but are known, the problem still can be handled by a slight modification of the procedure used for the common known variance case.

If the variances are equal from lot to lot, but the common variance is unknown, then a single-sample procedure cannot guarantee the experimenter's requirement. However, under this condition on the variances, a two-sample procedure can guarantee his requirement. The theory underlying the two-sample procedure is given in (2); tables necessary for its application are given in (5) and (6).

Finally, in the case of a common known variance of tensile strength from lot to lot, it is possible to use a sequential multiple decision procedure instead of the single-sample procedure. Both procedures guarantee the same requirement for a given goal and specification. The total sample size required by the sequential procedure varies from experiment to experiment. However, on the average the sequential procedure requires a total sample size per experiment which may be

substantially smaller than the fixed total sample size required by the single-sample procedure. The theory underlying the sequential procedure is given in (3) and (4).

References:

(1) Bechhofer, R.E.: "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 16-39.

(2) Bechhofer, R.E., Dunnett, C.W., and Sobel, M.: "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," Biometrika, Vol. 41 (1954), pp. 170-176.

(3) Bechhofer, R.E., Kiefer, J., and Sobel, M.: "A sequential multiple decision procedure for certain identification and ranking problems," (in preparation; to be submitted for publication to the Annals of Mathematical Statistics).

(4) Bechhofer, R.E. and Sobel, M.: "A sequential multiple decision procedure for certain ranking problems involving Koopman-Darmois populations," (in preparation; to be submitted for publication to the Journal of the American Statistical Association).

(5) Dunnett, C.W. and Sobel, M.: "A bivariate generalization of Student's t-distribution with tables for certain special cases," Biometrika, Vol. 41 (1954), pp. 153-169.

(6) Dunnett, C.W. and Sobel, M.: "Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution," (accepted for publication in Biometrika, June 1955).

(7) Wald, A.: Statistical Decision Functions, John Wiley & Sons, Inc., New York, 1950.

Table I

Table of $\sqrt{N}\lambda$ corresponding to various probabilities, to be used for designing experiments involving k normal distributions to decide which one has the largest (or smallest) average.

Specified Probability (P^*)	Number of Distributions					
	k=2	k=3	k=4	k=5	k=6	k=7
0.99	3.2900	2.6173	3.7970	3.9196	4.0121	4.0861
0.98	2.9045	3.2533	3.4432	3.5722	3.6692	3.7466
0.97	2.6598	3.0232	3.2198	3.3529	3.4528	3.5324
0.96	2.4759	2.8504	3.0522	3.1885	3.2906	3.3719
0.95	2.3262	2.7101	2.9162	3.0552	3.1591	3.2417
0.94	2.1988	2.5909	2.8007	2.9419	3.0474	3.1311
0.93	2.0871	2.4865	2.6996	2.8428	2.9496	3.0344
0.92	1.9871	2.3931	2.6092	2.7542	2.8023	2.9479
0.91	1.8961	2.3082	2.5271	2.6737	2.7829	2.8694
0.90	1.8124	2.2302	2.4516	2.5997	2.7100	2.7972
0.88	1.6617	2.0899	2.3159	2.4668	2.5789	2.6676
0.86	1.5278	1.9655	2.1956	2.3489	2.4627	2.5527
0.84	1.4064	1.8527	2.0867	2.2423	2.3576	2.4486
0.82	1.2945	1.7490	1.9865	2.1441	2.2609	2.3530
0.80	1.1902	1.6524	1.8932	2.0528	2.1709	2.2639
0.75	0.9539	1.4338	1.6822	1.8463	1.9674	2.0626
0.70	0.7416	1.2380	1.4933	1.6614	1.7852	1.8824
0.65	0.5449	1.0568	1.3186	1.4905	1.6168	1.7159
0.60	0.3583	0.8852	1.1532	1.3287	1.4575	1.5583
0.55	0.1777	0.7194	0.9936	1.1726	1.3037	1.4062
0.50	0.0000	0.5565	0.8368	1.0193	1.1526	1.2568
0.45		0.3939	0.6803	0.8662	1.0019	1.1078
0.40		0.2289	0.5215	0.7111	0.8491	0.9567
0.35		0.0585	0.3578	0.5510	0.6915	0.8008
0.30			0.1855	0.3827	0.5257	0.6369
0.25			0.0000	0.2014	0.3472	0.4604
0.20				0.0000	0.1489	0.2643
0.15						0.0364
0.10						
0.05						

Note: This table has been abstracted from Table I in Reference (1).

REDESIGNING FOR PRODUCTION

Harry G. Romig
International Telemeter Corporation

1. Introduction - Factors in the Development of New Products

Quality evaluation at all stages of the operations required for the production of a given item is one of the duties of a good Quality Engineering organization. From the inception of the concept underlying a new product, through research, development, production engineering, procurement, qualification tests, manufacturing, sales, costing, inspection, performance tests and shipping, pertinent data must be obtained and analyzed in order to make certain that the desired objectives of the design are attained.

In this seemingly endless chain, one of the most important links is that of production design. In actuality it is a redesign of the model provided by research and development, which, they believe, is a product that will perform the functions desired. In obtaining the original design the engineers have conducted critical tests to evaluate the performance of the unit and they now feel that it is ready for production. The various models and prototypes serve as examples of what can be achieved, but these as currently designed may prove too costly for economical production in the plant. Hence it is necessary to redesign for production taking into consideration all the factors that lead to the consummation of the intent of the original design, plus features that make it possible not only to produce it economically but also to provide cheap and efficient maintenance. Features of such production redesigns and the impact of quality engineering on such designs are discussed herein.

2. Relationship between Production Design and Quality Engineering

In designing for production there have been two schools of thought. In crash programs in particular, it has been felt that it was necessary to redesign on the production line while units and components are still in the development stage. The bare outline of the unit is made and is then supplemented by engineering changes that may be necessary in order that it may perform its function. The production line is itself used as the design laboratory. Production engineering is obtained through a liaison between the development engineers and the production supervisors. For a considerable time until the design is frozen, usually no two consecutive units are the same. The second unit incorporates the features of the first plus modifications that have been formulated during the production processes. The exponents of this method believe that such operations speed up the production process so that units for test come off the line much sooner. Their hope is that soon satisfactory and reliable units will be rolling off the line.

The second procedure is based on the philosophy that each group completes its work before turning it over to the next group. Research completes its studies and such bread-board models as it has used in its work. The development group takes these basic ideas and develops models and prototypes until it has a working unit that will stand its tests. It may carry out environmental and life tests on such units in their various stages of development to assure itself that it has the proper reliability and characteristics for a complete working model. This work is then turned over to the product engineer to redesign in line with the best and most economical manufacturing practices, working closely with the production engineer in charge of tooling and with the development engin-

eers so that the basic concepts of both research and development remain intact. Pilot runs are then made and units from such first production are given qualification and field tests until the unit has been proven. When the final results of these trials have been completed and the engineering changes resulting from the evaluation and analysis of them have been incorporated in the production design and checked adequately, then, and only then, is the product released to manufacturing.

The first method was used in the majority of plants during the past war and for a time thereafter. However, now the pendulum has swung the other way and in most instances companies, spurred on by the Military, are using the second method. For most types of products in the long run, reliable units are obtained more quickly from the second method. The product has more of the inherent "bugs" in the design removed prior to production. Also under normal circumstances, both the unit cost of manufacturing and maintenance are reduced.

Quality Engineering has great difficulty in handling products under the first method, particularly those phases related to inspection and test. Blueprints are constantly being changed. Standards of quality are usually unknown. The pressure is placed on all groups to get out something, no matter how crude and no matter how it differs from the available blueprints, and have units available for test as quickly as possible. Engineering changes are being made perpetually and it is difficult to evaluate the effect of such changes on the System as a whole. Insufficient time is usually allowed for lead time and for the obtaining of good components. Substitute materials are used when often undesirable. Quality Engineering must keep in constant touch with the engineers, procurement agencies, sub-contractors and manufacturing. New testing equipment must be designed and made. Since it is possible to have units, no one of which is like another, more inspection and test are required to maintain some control over quality. It is difficult to set standards and maintain them. Also maintenance costs are kept high since much more inventory must be kept to adequately service all the varied models that are in the field.

The second method is much more orderly and permits the careful analysis of results and evaluation of designs to determine theoretically and experimentally the best and most reliable for each function. Test equipment can be produced that determines with a minimum of tests whether the components, units, sub-assembles, assemblies and the System as a whole will perform as desired. Engineering can make the changes that seem best and qualify by adequate environmental and performance tests that design which seems best. The qualification tests under this procedure can be completed prior to production and the majority of the areas of trouble can be eliminated. Quality Engineering plays an important role in these operations by assisting in the proper design of experiments, keeping controls over the quality of materials and components obtained from outside sources, evaluating the tooling and checking such tooling against their engineering specifications. It may assist in setting up mathematical models for study and controls in line with the best procedures in Operations Research. Through analysis of laboratory, production and field data, testing equipment may be evaluated so as to obtain the proper relationships between test set errors and tolerances. With close teamwork between Production Engineering and Quality Engineering, the stage is properly set for releasing completed specifications for full-scale production that permit proper scheduling, lead time and delivery of reliable finished units.

3. Preliminary Analyses of Requirements

In many cases for large systems, contracts, either commercial or military, are prepared by both parties, seller and purchaser. These contracts need to be studied carefully from all angles. Quality Engineering must study those parts that relate to performance, tests, inspections, design requirements, special design checks, qualification and production tests and measurements and all features related to the final evaluation of the quality of the finished product. Such an analysis makes Quality Engineering conversant with the demands of the consumer and what features need to be checked on an 100% basis, which should be checked on a sampling basis, which are destructive, which are semi-destructive and which are not important but are merely window dressing. Objections to paragraphs which will impair the efficiency and resultant quality of the operation should be voiced prior to the signing of the contract so that the contract is reasonable, just, and to the best interests of all concerned.

A study should be made of laboratory requirements, manufacturing requirements, and field requirements. In many cases field testing may not be too good so that the errors reported back from outside purchasers may not be the ones which should receive the ultimate of attention. Laboratory facilities, within and outside, should be checked. The specifications offered by Production Engineering should be checked for difficulty of administering and the results of all such analyses, both numerical and empirical, as well as theoretical, should be tersely presented in tabular, graphical or pictorial form, together with concise statements covering the study.

When all groups have covered all phases of the requirements, material demands, special requirements and tests, legal aspects, schedules, guarantees, demands of maintenance, etc., these should be combined into an engineering report for consideration by Management prior to signing, not afterwards. Such a study will indicate what changes must be made in either the contract, design, or both, in order to obtain a clean-cut document which outlines clearly all requirements and the demands placed on all groups to fulfill such a contract.

4. Design of Production Experiments

The research and development engineers have set up an ideal for this product which they hope to attain through its production. They expect the ultimate of performance. It will be necessary to compromise some of these ideals in order to make a workable unit in manufacturing. Various designs for single parts have been made by research and development. Production engineering finds it impossible to make some of the weird shapes and objects envisaged by research and made with extreme difficulty by development without excessive costs. One combination of factors using very expensive and scarce materials may provide an ideal design. A cheaper replica or modification may work almost as well. Quality Engineering uses its standard analysis procedures to set up the type of experiment that will determine the variation in the results between the ideal and its cheaper possible replacement. Tests under all the variable conditions to be met should be arranged in the form of standard designs, experimental such as the Factorial, Block, Latin Square and Lattice, in such manner as to render with a minimum number of units in sample, a reasonably exact evaluation of the different designs in terms of their critical characteristics.

Before carrying out an expensive experiment, Quality Engineering

should obtain from research, development and production engineering a composite outline of the problem. The rules to be followed are: State the objectives; list the different conditions; note all possible solutions or designs known; reduce the number of potential designs to a minimum where some are not practical or are known to be too costly. For example, one design makes use of vacuum tubes while another uses transistors. Which is best? Tabulate estimates of initial costs, assembly costs, performance, maintenance costs, pro-rated loss from potential failures and all other pertinent factors. All must be noted and placed in their proper perspective for making the best possible experiment to prove as conclusively as possible the selection of the best and most feasible design for the task to be performed by the unit.

Quality Engineering sets the standards and then evaluates the actual performance of the different designs to assist in the proper evaluation of each fundamental design. IBM equipment or other automatic equipment should be used in such analyses where justified. This permits the use of larger samples at the same analysis cost and minimizes the errors in test. Proper designs result in simpler units with less failures and with better performance data. Critical designs of experiments pay off. Quality Engineering provides the method for lay-out and analysis and adds strategic information where, when needed, it may be utilized in freezing the design for manufacturing for a number of production units covering at least one block of serial numbers.

5. Tolerances -- Single and Composite

In redesigning a unit, the setting of proper tolerances is necessary. To properly evaluate those tolerances originally established by the development engineers, it is necessary to analyze the resultant effect of each critical tolerance. For individual dimensions which do not affect others, such an analysis is relatively simple. In this connection is it really necessary to hold such a strict tolerance on certain dimensions in order to assure proper operation of the unit? There are many open dimensions which have tolerances of ± .010 just because that is the general rule in drafting. For example, where a nominal dimension is given at say 3.000", and the note at the bottom of the blueprint states that all dimensions to .000 carry automatically the plus or minus ten mils noted above, a length is nonconforming if less than 2.990" or greater than 3.010" whereas parts between 2.900" and 3.100" may be usable. The careless use of such a rule results in needless scrapped parts. In many cases this rule is desirable, but in many at least thirty mils should be allowed to permit the use of commercial stock sizes thus reducing costs.

The principle above is also illustrated in the case of the establishment of the location of holes on a chassis for mounted electronic parts. Their exact positioning makes very little difference, since they are linked by flexible wires and not linked mechanically. All that is required is that they have a close enough relation between the parts so that they may be connected properly. Those using fractional tolerances state that ± 1/32" could be allowed with respect to such positioning. Others claim it could even be larger in many cases. However, usually the ten mil rule is applied as a general rule and many satisfactory chassis from a use point of view are relegated to the scrap pile.

Where composite tolerances are involved and the tolerances tend to add, it is necessary to consider how these tolerances interact. There have appeared many articles in this connection, some in ASTM publications,

others in IRE transactions Bartky and Ettinger more than a decade ago presented a discussion of rectangular, triangular and normal distributions, while Eugene Goddess covered this problem in connection with electronic components in a paper given before the IRE in New York. These various papers are using the Theory of the Propagation of Errors. The simplest exposition of this theory is presented by Deming in Reference (1). Reference (2) presents practical applications.

Theory considers that errors add up in a random fashion according to the Root-Mean-Square (R.M.S.) Law. If one dimension has a tolerance of 5 mils and four dimensions having this same tolerance are to be piled together, the resultant tolerance should not be four times 5 mils or 20 mils but should be the square root of four or two times 5 mils or 10 mils. This will only hold providing the parts are obtained from all parts of the parent populations in a purely random fashion and also are mated in a random fashion. Otherwise it may break down badly.

In practice, it is found that parts packed in boxes tend to closely approximate each other in their various characteristics. The level of the various boxes may be quite different due to random selection of boxes for shipment. Their composition tends to vary with time as machines turning out parts tend gradually to shift in their levels from hour to hour. The consumer obtains only a "chunk" from the total distribution. Hence chunks are to be matched together with other partial sections or chunks from other distributions. The resultant tolerances may tend to be on one side at one time or on the other side for other component parts. Allowances must be made in setting over-all tolerances to care for this practical situation. Some engineers add tolerances and take arbitrarily 0.7 times such sum. This is often sufficient allowance. In electronic components the factor 0.8 has been used and found to fit the data much better than the R.M.S. rule. In evaluating composite tolerances, the Quality Engineering group can be of great assistance to the Production Engineers in determining from actual data what are reasonable over-all tolerances. When individual tolerances are already established, the over-all tolerance can be determined in accordance with the above rules, but the values thus obtained must be tempered by service demands. The reverse problem is usually met with in practice. What should be the relationship between individual tolerances to achieve the desired over-all tolerance most economically? In one instance just by a re-evaluation of about 32 additive tolerances, parts that had to be carefully matched could be assembled on the line and made to operate quickly by a quick run-in whereas previously parts had to be carefully screened to secure units that would fit together. Such redesigns are necessary to obtain quality products at minimum costs.

Emphasis must be placed on the fact that under the extreme environmental conditions which are placed on designs today, large tolerances, not tight tolerances are necessary to care for extreme changes. Precision parts may not work. Parts with liberal allowances will. Redesigns for production must take account of this fact and provide maximum permissible tolerances wherever possible. This makes it possible to give a little more tolerance to those critical dimensions that are difficult to maintain and tighten up on dimensions whose tolerances are easy to maintain. In summary, the possible rules may be expressed as follows:

$$\text{R.M.S.:} \quad T = T_1^2 + T_2^2 + \ldots + T_m^2 = \Sigma T_i^2 ,$$

Empirical: $T = k(T_1 + T_2 + \ldots + T_m) = Tk\Sigma T_i$,

where m = Number of components contributing to the over-all tolerance,
k = constant. For $m = 10$ or less, $k = 0.7$ is sometimes used for mechanical tolerances and some electrical tolerances, and $k = 0.8$ for some electrical tolerances.

6. Selection of Standards

All departments in a company are interested in the establishment of standards for their work. Those standards which affect the company most vitally are the Quality Standards established by Quality Engineering. Sales are anxious to know how their product compares with that of their competitor. What are the sales points that can be stressed in order to show superiority over a competitor's similar product? Engineering considerations, information received from field trials, and complaints form a basis for the establishment of standards that truly reflect the demands of service and the ability of the company itself to design and produce a satisfactory product.

Quality Engineering works closely with the other groups in establishing economical standards for any new product being redesigned under Production Engineering. Having analyzed data covering similar products and evaluated those cases which have little bearing on quality as well as those which truly reflect consumer demands, appearance factors, life characteristics, reliability, ease of maintenance, particularly the ability to readily obtain interchangeable parts when units wear out in service, standards of quality are incorporated in the design as much as possible. A good design is usually simple and provides easy access to internal parts. In electronic equipment there is a strong tendency to make a rat's nest of the wires rather than to make small interchangeable sub-assemblies that not only are neat in appearance but also make the maintenance problem much easier.

Consideration is given to costs as well as to materials and consumer preferences. In practice, if possible, a rating system is used for the final evaluation of quality. When carried to its ultimate goal, the use of either demerit rating systems or some type of Index makes it possible to compare various products that are similar in nature, also piece-parts, sub-assemblies, assemblies, and even shops with each other and with the established standards. Such standards must be reviewed continually, but when well established, they provide a guide in the designing of new products.

7. Engineering Changes -- Block System of Controls

In the early stages of design it is impossible to determine what changes will be made by the various engineers working on various phases of the project. In some cases two or more groups may be working in different areas to obtain designs for the same final purpose. Hence engineering changes are being made continually. After the development stage, however, it should be possible to redesign the product so that in the future except for basic changes, there will be a minimum of engineering changes made. This is true under the second approach where the qualification tests must be satisfied before release to production is authorized.

The problem of inventory, maintenance, production controls,

scheduling, purchasing component parts, sub-contracting, all these and more are tied up in the methods used to handle engineering changes. If no controls are exercised, then no one is certain that the units currently produced are being made to the most recent requirements or not. By proper controls over blueprints and specifications and their distribution, many of these difficulties can be readily overcome. It is not easy to keep the engineering drawings up-to-date when the area of knowledge is continually shifting. Those controls that are necessary are exercised by the Quality Control group in most companies who often have jurisdiction over the dissemination of information covering all engineering changes as well as blueprints.

The Military, in particular, are interested in the controls that are exercised over all requirements and, in particular, over the incorporation of engineering changes in the product as quickly as possible. This saves them the possibility of having to authorize modifications later, if the changes made justify making the changes retroactive. Means must be devised to maintain controls and adequate records of all changes. Also all interested parties must know of all changes as far as possible at the same time.

The controls that seem to work the best use the "Block System" of controls. This freezes a design for units serialized from one specified number to another. All units covered by such serial numbers are made in accordance with the design specifications in effect at the time of the freeze. In the meantime the other engineering changes and modifications are studied by all concerned to determine their effectiveness on the System as a whole. These are then assembled as a unit and where approval of the Military is required, such is obtained. These changes then are placed en bloc in effect covering another series of units. This is probably the best control that can be exercised. Quality Engineering watches these changes carefully and all inspection and tests are made in accordance with the specifications in effect at the time of the freeze involving that particular series of units.

In other instances the Block System is applied to sub-assemblies rather than to Systems as a whole. This appeals to the engineers more as the development of the various parts often lag behind each other. It has merit but also many dislike it from a maintenance point of view as there will be very few units that are similar in all respects. The effectiveness of the changes will differ from sub-assembly to sub-assembly so that the maintenance engineer never is certain what parts must be kept on hand unless a record of all effective points of changes are kept on file for immediate use and for ordering purposes.

Probably some compromise must be adopted. For very complex Systems, the Block System for Major Sub-Assemblies might be applied. Otherwise the Block System should be used for each product.

8. <u>Qualification and Production Tests</u>

In the early stages of design the engineers are expected to conduct their own tests and also must devise testing equipment to determine the suitability of their designs. These test equipments will later prove useful in production if they are properly designed and maintained. In Production Engineering, after each redesign is made it is absolutely necessary to carry out the complete qualification tests that may be a part of the initial contract in order that it can be determined whether the production redesign is truly satisfactory. Quality Engineering should

not take this function away from Production Engineering but should conduct a spot check of certain tests and also watch the carrying forward of these qualification tests. In many instances such tests involve not only the standard checks usually required but also some severe environmental and vibration checks that are required under the heavy demands that are now being placed on products which may possibly be used by the Military.

The tests that are the responsibilities of Quality Engineering are those that are designated as the Production tests. These are often spelled out in the specification in detail. Some groups and engineers desire that these give exactly the number of units that must be measured as well as the methods of test that must be applied. However this is frowned upon by the Quality Engineer as it does not permit him to increase the tests on the poor supplier and take such tests on a periodic basis for the better supplier that has true control over his productive processes.

In line with this thinking it is necessary for Quality Engineering to set up a system for analyzing such qualification and production test results in a form that can be presented to Management. This will provide one of the best checks on the quality performance of not only the unit but also the various vendors and sub-contractors that are involved in any particular contract. Such evaluation will give Management a different view of the entire project. The weaknesses will show up quickly and it is then possible to make the necessary corrective action so that the outgoing product will truly reflect the desires of the company to produce a quality product as economically as possible.

9. Making Engineering Teamwork Pay

The presentation has not been technical in nature as my desire has been to discuss the problem in its entirety rather than the little bits and pieces that require detailed technical information in order to obtain the best possible solutions. There are many excellent texts that cover the details and also current magazines are filled with new ideas covering production problems. These need to be carefully reviewed in order to apply these ideas in places where they best fit.

Because of the rapidly changing picture, the development of new materials and techniques, it is no longer a one man program. The day of the single genius is over except in a few fields. The demand is for teams of individuals with diversified skills and knowledge that can work together collectively and produce systems rather than pieces. The thought is that such teams will develop new ideas to their ultimate conclusions and that such products as arise will be far superior to those developed by the lone scientist and inventor.

This concept follows along the lines of Operations Research and does provide a basic pattern for developing a product that embodies all the latest changes and reliability features at the most economical costs consistent with performance. Each group has done its part and has contributed to the whole. The research group have singly or collectively developed the new idea with all the logistics connected thereto. The development group have determined its feasibility and have made such models as are necessary to develop it to a stage where it will perform the functions assigned. The product engineer has made up a new design that is possible to make in production and has performed such qualification tests as are necessary to determine that the new design will

meet the demands of the prior groups. Then in manufacturing the production engineer has devised the necessary tooling to carry on production effectively.

Throughout this complete program, Quality Engineering assists where desirable and makes such additional tests and analyses as may be required to prove the effectiveness of the resultant product at all stages of its growth. As a result when it is time to produce the finished product on a grand scale it has the necessary knowledge and testing equipment to evaluate the finished product, set the best possible commercial standards and assist all in producing this product as economically as possible. It keeps scrap down to a minimum and also reduces the possibility of customer complaints. Such teamwork pays off and produces the best possible products at minimum costs.

10. References

(1) W. Edwards Deming, "Statistical Adjustment of Data," John Wiley & Sons, Inc., New York, 1948, Chapter III, The Propagation of Error.
(2) Gardner M. Hailes, "Some Statistical Principles of Tolerances," Industrial Quality Control, Vol. VII, No. 6, May, 1951, pp. 77-81.
(3) Donald Statler Villars, "Statistical Design and Analysis of Experiments For Development Research," Wm. C. Brown Co., Dubuque, Iowa, 1951. (Practical Text for Analysis of Variance.)
(4) Leonard A. Seder, "Diagnosis With Diagrams - Part II, Case Study - Graphic Analysis of Variance," Industrial Quality Control, Vol. VI, No. 5, March, 1950, pp. 7-11.
(5) Irving W. Burr, "A New Method for Approving a Machine or Process Setting - Part II," Industrial Quality Control, Vol. VI, No. 2, September, 1949, pp. 15-19.
(6) Raymond Villers, "The Dynamics of Industrial Management," Funk & Wagnalls Co., New York, 1954.

THE BENDIX RADIO VENDOR QUALITY RATING SYSTEM

H. C. Newton and W. A. MacCrehan
Bendix Radio Division of Bendix Aviation Corp.

I sometimes feel that our use of the phrase "Vendor Quality Rating System" is a mistake. It implies--I'm afraid--that we have devised a technique which is nothing more than a coldly mathematical procedure for determining which of our suppliers are good and which are bad.

Now, it is true that the Vendor Quality Rating System does provide us with numerical ratings of the quality of the products shipped to us by our suppliers. And, of course, this information is valuable. But I want to emphasize the fact that vendor rating _per se_ is only one phase of the program.

Consider, for a moment, the volume of material that comes into the Bendix Radio Receiving Department every month: Each month, our plant receives an average of 10,000 shipments of materials and parts, sometimes comprising as many as 20 million units. The names of more than 700 firms are carried on our list of active vendors. They supply us with a variety of items ranging from printed labels to radar antennas. With these figures in mind, it becomes obvious that the quality equipment produced in our plant is--in a very real sense--the result of a cooperative effort of Bendix Radio _and_ its suppliers. Indeed, this situation is common throughout industry today. To paraphrase a very profound observation: In this age of specialized manufacturing, no company is "an island entire of itself".

Our Vendor Quality Rating System was designed to take cognizance of this interdependence between Bendix Radio and the vendor. Experience had taught us that the unimaginative "accept-reject" procedure--all too common in industry--was wasteful and ineffective. It took no thought of the causes of defective material coming into our Receiving Department. We reached the inevitable conclusion that we must get the vendor on our side . . . that we must devise a system which would enable us to work together on the problem of producing Quality. We learned--in short--that we must begin to take the "broad view" in our relations with our vendors. The culmination of this thinking was the Vendor Quality Rating System.

Having recognized the problem, and with a fairly definite idea of what we wanted to accomplish, we initiated a study of the data available on purchased parts and materials. In this, we got valuable help from our Purchasing Department. I might say here that, without the cooperation of the Purchasing Department, we could not have achieved success in the program.

Perhaps the first thing we discovered in our study was our need for an accurate, continuing picture of the quality of purchased products. In order to accomplish our aims, we had to have some means of evaluating quality trouble when it occurred. The statistical methods used to accomplish this are discussed elsewhere in this paper.

Further study revealed that, many times, so-called "quality problems" were not quality problems at all, but misunderstandings, difficulties in liaison, misinterpretation of specifications, etc. This pointed to the need for closer cooperation between the vendor and Bendix Radio;

The focal point of the Vendor Quality Rating System is the Vendor Quality Rating Board. Mounted on this board are individual Rating cards for each of Bendix Radio's more than 700 active vendors.

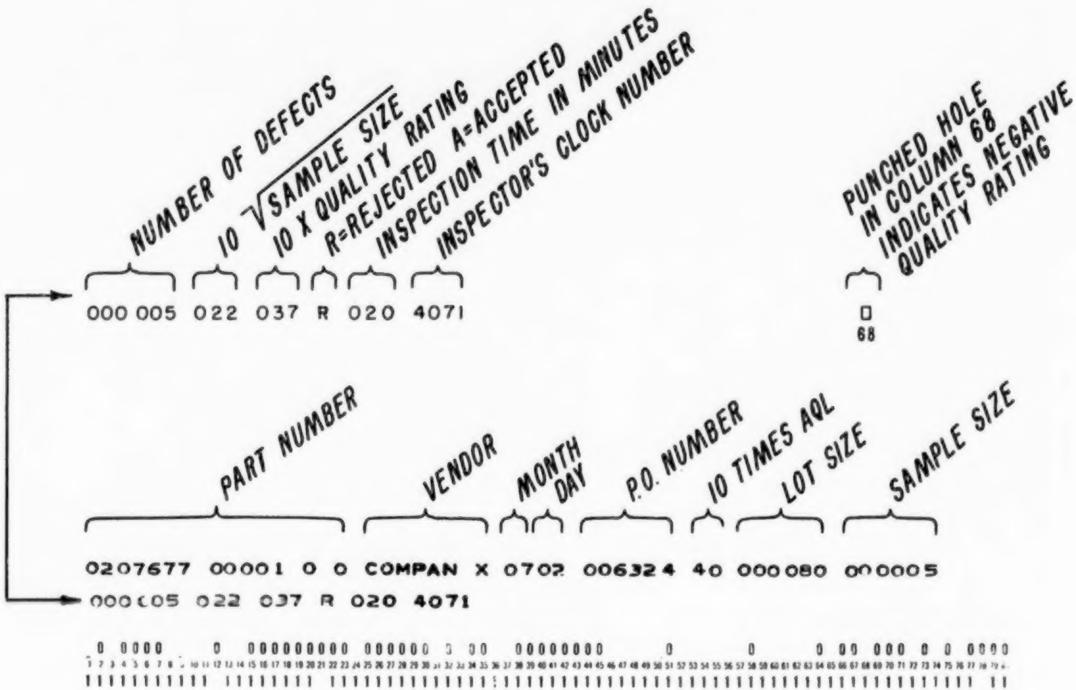
specifically, between the vendor and the Bendix people assigned the responsibility for maintaining our quality standards. Events proved that this "person-to-person" approach was highly effective.

Without exception, our vendors have given us their cooperation. From the beginning we made it clear that we were asking them to join in a mutual effort to solve the problems that concerned us all. We took concrete steps to assure the vendors that our attitude in the matter was sincere. As a result, our vendors have responded with enthusiasm. In fact, they have volunteered many suggestions not only for the improvement of their own product quality, but to increase the effectiveness of the overall Vendor Quality Rating System. The vendors have found that they benefit not only in their dealings with Bendix Radio but with their other customers as well. We like to feel that the effect of this program will eventually be felt throughout our industry. We know that it has introduced a new concept of vendor evaluation in our own Division.

I have no doubt that--just about now--some very practical-minded person in the audience is saying to himself, "This is all very well, but what does it mean in dollars and cents?" The stock answer to this question is the fact that any improvement in the quality of a product or of the parts and materials that make up a product places the manufacturer in a better competitive position . . . and that is a valid answer; but--in the case of the Vendor Quality Rating System--I can be quite a bit more specific:

For many years, Bendix Radio Division has maintained an extensive Field Inspection program. At times, we have had as many as 25 field inspectors in our vendors' plants acting as a sort of advance guard in the effort to assure that all purchased parts met Bendix Radio quality standards. This was a necessary program, but an expensive one. In the year 1953, for instance, Field Inspection represented an expenditure of 38,872 man-hours. However, in the three-year period since its inauguration, the Vendor Quality Rating System has so improved the overall quality picture insofar as purchased parts are concerned that we have been able to reduce the Field Inspection operation by approximately fifty per cent . . . and this process of reduction is still going on. Considering that the Vendor Quality Rating System involves not more than 6200 man-hours a year to operate, and comparing this with the figure given above for an average year of Field Inspection, it can be seen that, in this one field alone, the Vendor Quality Rating System has effected considerable savings.

The Vendor Quality Rating System has also enabled us to institute reduced inspection plans on many purchased items processed in our Receiving Inspection Department. After the initial "bugs" had been worked out of the System and we began to see its effects, it became evident that the improved quality of many of the items supplied to us by our vendors would permit us to lighten the burden of Receiving Inspection. For the past two years, we have had a reduced inspection plan in operation. I must admit that we took this step very hesitantly. By means of our Quality Audit program, we kept a critical eye on the quality of the units containing parts affected by the reduced inspection plan. However, the results to date have been so gratifying that plans are now being prepared which will lead to further reduction in our Receiving Inspection. Incidentally, all of our reduced inspection programs are derived from the plans contained in MIL Standard 105.



The pertinent information for each lot received is key-punched into a Lot Rating Card, as shown here.

In addition to the benefits and savings already mentioned, the Vendor Quality Rating System has also done a great deal to improve the record-keeping function of our Quality Control Department. Those of you who are closely associated with an industrial quality control facility know that the collection, compilation, and recording of data comprises a large percentage of the Quality Control activity. This is due not only to the nature of the Quality Control function, but to the fact that there are specific requirements under military contract calling for the maintenance of extensive records. This is particularly true where reduced inspection plans are in operation.

With standard clerical methods, the record-keeping operations of a large Quality Control department are costly, and often inefficient. Moreover, the accuracy and availability of the information leaves a great deal to be desired. Vendor Quality Rating went far to solve this problem for Bendix Radio Quality Control. For it was the success of punched-card-machine procedures in compiling and recording the data specifically required for Vendor Quality Rating that encouraged us to extend the use of this same punched-card technique to the many other record-keeping chores of the Department. As a result, we have decreased the quantity of our paperwork, while increasing the accuracy and availability of our Quality Control data.

In my view, the success of the Vendor Quality Rating System provides practical proof of two very important facts: Number one is the fact that the technique of statistical quality control--when applied in a careful, common-sense manner--is a valuable industrial tool. And, number two is the fact that <u>when people get together face to face</u>, with all the factors in a problem before them and with a spirit of cooperation present, that problem is going to be solved.

I have discussed the "philosophy" behind our Vendor Quality Rating System and some of the benefits that we have derived from it. I'd like to tell you something about the actual workings of the System. The Vendor Quality Rating System consists of three principal phases:

1. Compiling data
2. Providing a continuous history of individual vendor quality
3. Cooperating with the vendor to solve the quality problem

The first phase is simplified by the fact that inspectors and testers from the Quality Control Department are strategically placed and easily available for the collection of data. The presence of well-established sampling plans further simplifies the problem by reducing the volume and increasing the significance of the data collected. But, even with these considerable advantages, the data collected is so complex and so extensive that it cannot be used as is for our purposes. The solution to this problem is our use of punched-card techniques. The various automatic punched-card machines proved to be ideal for the task of reducing our test and inspection data to manageable proportions.

The second phase of the System--that is, the actual quality rating--appears in the form of a large rack some six feet high and thirteen feet long. This rack contains hundreds of individual Vendor Rating cards . . . one for each of our active vendors. Each card carries a graph of inspection results, and has a colored tab attached to the upper left-hand corner, in the visible margin, alongside the vendor's name. Plotted on the graph is the vendor's month-by-month quality rating. With 100 as

IBM operator running off a tabulation of vendor quality information.

the top rating, 90 to 100 is considered top quality, and the vendor who stays in this category earns a gold tab on his Quality Rating card. A green tab indicates an acceptable rating of 50 to 90. For a rating of less than 50, the vendor gets a yellow tab, indicating an unacceptable level of quality.

Any vendor whose products fall below the acceptable level for three months running, gets a red tab, indicating that corrective action must be taken immediately.

The information which appears on the Vendor Quality Rating chart is secured with the help of the complicated calculating machines in the company's extensive IBM section. The basic information is provided by the Receiving Inspection Department. People in this department inspect each incoming shipment or "lot". Of course this does not apply to shipments received from vendors covered by our reduced inspection plans.

Virtually all inspection in the Receiving Department follows a technique of random sampling derived from the Government specification on sampling (MIL-STD-105A). The inspectors prepare individual forms for each lot inspected, and at the end of the day an operator key-punches the information from these forms into individual lot cards. These cards contain the following information:

1. Vendor's name
2. Date of inspection
3. Part number
4. Purchase order number
5. Lot size
6. Sample size
7. Number of defects
8. Acceptable Quality Level
9. Whether shipment is accepted or rejected
   (When a lot is rejected, a supplementary card containing the cause of rejection is punched.)

Acceptable Quality Level, or AQL, is a standard term in quality control. Where service contracts are concerned, the AQL is fixed by the services involved.

The next step is the assigning of a quality rating to each lot. It is necessary that this rating be in the form of a simple, easily handled numeral. This problem was solved through the use of a test of significance formula. For the purpose of rating a single lot, the following statistic was computed:

$$t = \frac{p - p'}{\sigma_{p'}}$$

where

p = the percent defective of the sample quantity inspected

p' = the AQL percent

$\sigma_{p'}$ = the standard deviation of p'

$$\sigma_{p'} = \sqrt{\frac{p'(100-p')}{n}}$$

n = the sample size

This provided an equitable method for rating any one lot; however, the form of the answer was not satisfactory. It was necessary that the rating be presented on an easily understood scale. Moreover, it was felt that the concept of a "t" value might be foreign to the personnel of the Purchasing Department and the vendor. It was decided, therefore, to change the scale of the rating as follows:

$$\text{LOT RATING (L.R.)} = 70 - 10t$$

Thus, any lot with a percent defective equal to its AQL would receive a rating of 70. Assuming a significance level corresponding to two sigma, significantly good lots would have a rating of 90 ($t = -2$) or more, and significantly bad lots would receive a rating of 50 ($t = +2$) or less. The final form of the rating was then

$$L.R. = 70 - 10\,\frac{P - P'}{\sqrt{\frac{P'(100-P')}{n}}}$$

After the system for rating a single lot was devised, it was necessary to consider: a) the interval to be covered by the rating, b) how to take into account the fact that there are varying numbers of lots supplied by each vendor during a given interval and, c) whether the results were to be cumulative. Moreover, it was necessary that the Vendor Quality Rating System be adaptable to all types of vendors. It must be equally applicable to the large supplier and to the vendor who makes small and infrequent shipments . . . to the vendor who supplies many different kinds of items as well as to the vendor who supplies one or two types of items; and the ratings must be in the form of values that are readily comparable. It was decided to rate all vendors once each month on the basis of all shipments received during that time. These results would not be cumulative but, rather, each vendor would have a chance for a fresh start each month.

It was felt that, for vendors supplying lots of equal size to the same AQL, the greatest significance should be given to the vendor supplying the most lots within the given period of time. Our conclusion was that the proper approach was to compute, for each vendor, the average "t" value for all lots received that month. The significance of $\bar{t}$ would then be tested, and this test would yield the final monthly rating for the vendor. The test is as follows:

$$t'' = \frac{\bar{t} - \bar{t}'}{\sigma_{\bar{t}'}} \text{, but } \bar{t}' = 0 \text{ and } \sigma_{\bar{t}'} = \frac{1}{\sqrt{N}}$$

$$\therefore\ t'' = (\bar{t})\sqrt{N} \quad \text{where } N = \text{number of lots}$$

$$\text{thus QUALITY RATING (Q.R.)} = 70 + t''$$

$$\text{Q.R.} = 70 + \left(\frac{\Sigma \text{L.R.}}{N} - 70\right)\sqrt{N}$$

In order to further clarify the method used for computing Vendor Quality Rating, let us consider a hypothetical vendor "X", who has shipped three lots of material to Bendix Radio during the past month. The following progression illustrates the method of determining, first, the rating of each lot, and then the vendor's Quality Rating for the month.

<u>Lot number 1</u>

Sample size (n) = 64

Number of defects = 1

Sample % defective (p) = 1.6%

AQL% (p') = 2.5%

Lot Rating (L. R.) = 70 - 10t

where $t = \frac{p - p'}{\sigma p'}$ and $\sigma p' = \sqrt{\frac{p' (100 - p')}{n}}$

It follows then:

$$\sigma p' = \sqrt{\frac{2.5\ (97.5)}{64}} = \sqrt{\frac{243.75}{64}} = \sqrt{3.81} = 1.95$$

$$\text{and } t = \frac{1.6 - 2.5}{1.95} = \frac{-.9}{1.95} = -.46$$

therefore, Lot Rating (L. R.) = 70 - 10 t

L. R. = 70 + 4.6 = 75

<u>Lot number 2</u>

n = 32

number of defects = 0

$p = 0\%$

$p' = 2.5\%$

$L.\ R. = 70 - 10t$

where $t = \frac{p - p'}{\sigma p'}$ and $\sigma p' = \sqrt{\frac{p'\ (100 - p')}{n}}$

It follows then:

$$\sigma p' = \sqrt{\frac{2.5\ (97.5)}{32}} = \sqrt{\frac{243.75}{32}} = 2.76$$

and $t = \frac{0 - 2.5}{2.76} = -.91$

therefore, $L.\ R. = 70 - 10t$

$L.\ R. = 70 + 9 = 79$

Lot number 3

$n = 128$

number of defects $= 2$

$p = 1.6\%$

$p' = 4.0\%$

$L.\ R. = 70 - 10t$

where $t = \frac{p - p'}{\sigma p'}$ and $\sigma p' = \sqrt{\frac{p'\ (100 - p')}{n}}$

It follows then:

$$\sigma p' = \sqrt{\frac{4.0\ (96.0)}{128}} = \sqrt{\frac{384}{128}} = 1.73$$

and $t = \frac{1.6 - 4.0}{1.73} = \frac{-2.4}{1.73} = -1.39$

therefore, $L.\ R. = 70 - 10t$

$L.\ R. = 70 + 14 = 84$

Monthly Quality Rating (Q. R.) - $70 + \left(\frac{\Sigma L.R.}{N} - 70\right) \sqrt{N}$

where N = the number of lots

$Q.\ R. = 70 + \left(\frac{238}{3} - 70\right)\ 1.73$

$$Q.\ R. = 70 + (9)\ 1.73$$

$$Q.\ R. = 86$$

With a Quality Rating of 86, vendor "X" would rate a green tab on his vendor rating card.

It's in phase three that the Vendor Quality Rating System pays off. When the need for corrective action on any particular purchased part appears, our Purchasing Department contacts the vendor and invites him to send a representative to our plant to discuss the problem. A meeting is arranged, attended by the purchasing agent concerned with the part involved, the Receiving Inspection supervisor, a Quality Control engineer, and the vendor's representative. The meeting is conducted as a completely informal "round-table" discussion. The object of the discussion is to bring to light the "whys and wherefores" of the defects in question. There is a great deal of give and take and it is not all sweetness and light. Oftentimes, the discussion reveals the plain fact that the parts or materials in question simply do not come up to Bendix Radio quality standards and they cannot be accepted. When this happens, an effort is made to help the vendor improve his own Quality Control procedures so that he will be able to produce the rejected material successfully. We have even gone so far as to assign a Bendix Radio Field Inspector to the vendor's plant to help with the problem.

But, on many occasions, it is discovered that the problem is the result of differing interpretations of prints or specifications, tolerance requirements that are unnecessarily close, use of poor gages, etc. This type of difficulty is quickly and easily remedied, and the vendor's representative returns to his plant armed with the information necessary to do the job. Where the problem is more difficult, a second meeting is arranged so that the progress made in finding a solution can be determined.

Certainly the Bendix Radio Vendor Quality Rating System is no panacea for all the ills relative to the control of quality in purchased parts. But it does represent an approach to the problem that we think is healthy and very promising. Installing the Vendor Quality Rating System has helped our vendors to help themselves . . . now, both we and they are reaping the benefits.

THE STANDARDIZATION OF RAW MATERIALS, PROCESSES AND PRODUCTS IN TEXTILE MANUFACTURING

Oliver P. Beckwith
Fabric Research Laboratories, Inc.

Without standards control of quality would be impossible. We would have nothing to compare production with to determine whether or not it was satisfactory. Setting standards, then, is the beginning of quality control.

The functions of specifications and standards are primarily:

1. To communicate to suppliers the purchaser's requirements for acceptable material.

2. To insure that processes are performed and products made as they were designed and costed.

3. To insure uniformity of quality - that the product is made the same from batch to batch and lot to lot.

4. To insure that the product meets the customer's specifications (for example the Armed Forces).

A specification for purchased material consists of at least two parts as follows:

A. Design Requirements: This defines the dimensions, physical and chemical properties and performance characteristics which the material shall have.

B. Methods of Test: This defines the analytical engineering procedures to be followed in conducting an individual test.

Sometimes purchased material specifications contain a third part on acceptance requirements. This defines the nature and amount of evidence considered necessary to establish that a material complies with the Design Requirements. Many authorities believe that statements on sampling and inspection should not be included in the purchase specification. A cogent reason for this is that the most economical sampling plan to use depends on the quality level of the supplier. Hence, a particular sampling plan included in a purchase specification may be satisfactory in dealing with a producer having poor control but would require unnecessary inspection and testing for the one having good control.

The preparation of the Design Requirements of a purchase specification must result, essentially, from a meeting of minds of producer and purchaser. The quality levels required by the purchaser in his processing and finished product must be melded with the quality levels and variations thereof that exist in the production of the material in question.

Frequently this is accomplished by using specifications prepared by the American Society for Testing Materials. (1) The A.S.T.M. Committees are composed of balanced groups of producers on the one hand and consumers and general interest members jointly, on the other hand. The deliberations of the committees preparing specifications bring about the meeting of minds of producers and consumers. The procedure by which a

proposed standard or specification becomes an official standard of the Society is a democratic one and all interested members have the opportunity to voice their views on the proposals and to vote to accept or reject them.

While setting standards is the beginning of quality control, the development of test methods is truly the beginning, when properties of material are to be assessed by quantitative measurement. A standard value for a property of a material has meaning only in reference to the test method employed. For example, the tensile strength determination of a textile will vary widely depending on the method of test employed and the type of testing equipment. The raveled strip method, the grab method, the use of constant rate of traverse testers (pendulum types) or constant rate of load testers (incline plane types) will affect the value of tensile strength. It is, therefore, necessary that the method of test be precisely defined in the standard or reference be made to a particular test method described in the standards of the American Society for Testing Materials or the American Association of Textile Chemists and Colorists.

Sound methods of test should give essentially the same results when buyer and seller test the same material. The development of such methods is not an easy matter. Considerable research and development on an industry-wide basis goes into the preparation of both A.S.T.M. and A.A.T.C.C. standards. A recommended practice for interlaboratory testing has been prepared by Committee D-13 on Textiles of A.S.T.M. (2). It describes the application of statistical principles in the planning of interlaboratory studies for test method development and the collection, analysis and interpretation of the data obtained in these studies. The techniques used are helpful in revealing the sources of variability and effecting improvements in the proposed method.

If Acceptance Requirements Requirements are to be included in a purchase specification they should specify:

1. A well defined and standardized technique for selecting a sample.

2. Who is to perform the sampling inspection - producer or purchaser.

3. Where the sampling is to be done - at point of manufacture or at customer's plant.

4. Whether a lot is to be randomly sampled or divided into stratas and randomly sampled.

5. The size of the sample to take for a given lot size. The number of tests to make on a single unit of a sample.

6. The action to take when a sample is tested for more than one characteristic and some but not all the requirements are met.

7. The statistical measures to be computed, e.g., average, standard deviation, range, et cetera.

8. Whether the average of the lot must meet the values given in the Design Requirements section; all units in the lot; 95% of the units, et cetera.

9. The action to take when the sample does not meet the acceptance criteria. There must be a clear definition of what rejection applies to.

10. A sampling plan based on a careful study of the process capability of producers in the industry.

The preparation of acceptance requirements necessitates the application of statistical quality control methods. Most testing and inspection of textiles must be done on a sample that is taken to represent a lot. 100% inspection is generally impossible because testing often destroys the material or the cost of 100% testing is prohibitive. In judging a lot by a sample there is always a risk of rejecting a good lot or accepting a bad lot. The use of statistical methods enables one to specify the risk of such occurrences for a given sample size and process quality level. When this is known the costs of testing can be balanced against the losses incurred from accepting a bad lot or rejecting a good one. The most economical sample size for maximum protection can then be specified.

Practices have been developed in A.S.T.M. Textile Standards and in Military Standards (3) which are helpful in the problem of writing acceptance requirements in specifications. The "Recommended Practice for Calculating Number of Tests to be Specified in Determining Average Quality of a Textile Material" (4) found in the textile standards of Committee D-13 treats the problem of specifying the sample size to take to assess lot quality with a desired precision and probability. The same problem, in different form, is covered in A.S.T.M. Textile Standard D1060-53T, Core Sampling of Wool in Packages (4). The statistical techniques used here are applicable to other baled or packaged materials, whether they be cotton, rayon, dyestuffs, soap, et cetera. Military Standard 105A, Sampling Procedures and Tables for Inspection by Attributes, describes sampling plans that can be readily adopted for such situations as the inspection of woven textiles.

Sampling plans in purchase specifications ought to be flexible enough to cope with the different situations met in practice. When one lot only of material is to be purchased and nothing is known about the process quality level of the producer, good sense would indicate a larger sample size than when testing successive large lots from one producer on which a long quality history has been obtained. The number of tests necessary on material from a producer having good control will be less than that from a producer having poor control. Provision should be made for these different situations in preparing the sampling plan.

A relatively recent development in industrial specifications and inspection is the concept of acceptable quality levels. This is recognition in the specification of the fact that in some products it is more economical for the purchaser to suffer a certain amount of defective material than to demand perfection in each lot. For example, in woven textiles a percentage of short rolls is always generated because of the running out of beams. To eliminate such rolls would mean the creation of considerable waste, therefore, perfection in roll length would mean added costs to the purchaser. When analysis shows that the added costs

incurred by the purchaser in using a certain percentage of short rolls are less than the increased costs charged for all full length rolls then it makes sense to accept a percentage of short rolls. This holds true in the inspection of woven textiles for appearance defects. The concept has probably been accepted in the textile trade since the beginning of mass production of textiles. However, by specifying the acceptable quality level and using statistical sampling plans the purchaser can economically assure himself that this level is not exceeded in the long run. The Quartermaster Inspection Service employs this principle in its acceptance inspection of textiles for the Armed Forces. Military Standard 105A and the Dodge-Romig Tables (5) are useful in developing sampling plans for acceptance inspection of textiles.

Many textile mills and garment plants consume large quantities of yarns, thread and fabric. Purchase specifications for such material are useful and practical. There has been some resistance to the use of specifications in these transactions, not only on the part of the vendor, but in the purchasing department of the buyer. The argument against their use is that they would increase the vendor's costs. Since this, presumably, would be passed on to the buyer, his purchasing agent is not keen on making a move that would apparently adversely affect purchasing performance.

No doubt the use of specifications with their provision for the rejection of substandard material would add to the costs of a vendor whose product was poorly controlled. In a competitive market such a vendor could not add these costs to his price. Those competitors who can meet the specification at the required price would get the business. Thus, the manufacturer having poor control is forced to improve his operation or else suffer reduced profit. When action is taken to improve quality by preventing defects, reduced costs are the end result. So the widespread use of purchase specifications opens the way to better quality of material at reduced cost rather than increased costs.

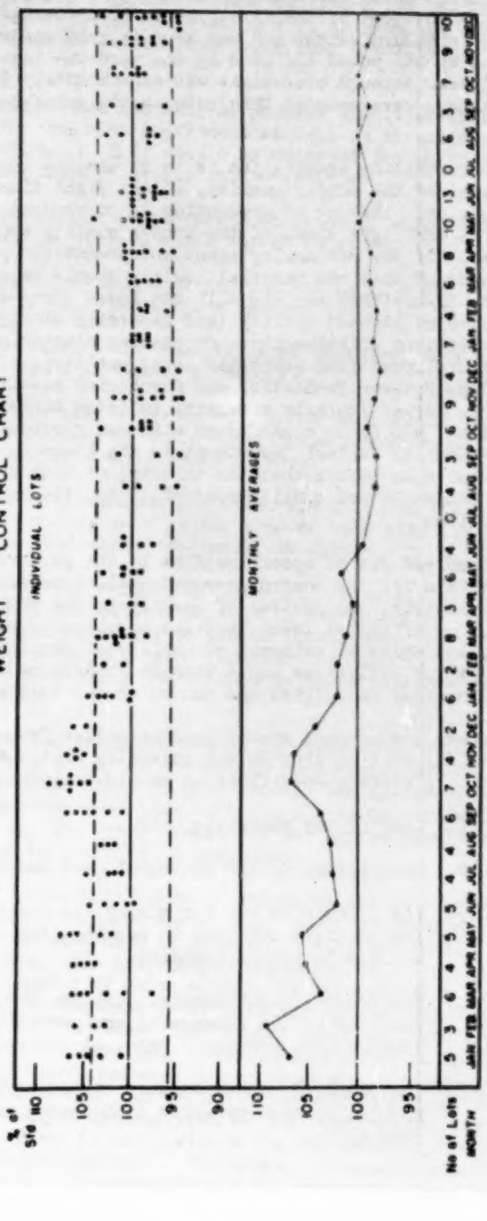
The use of sound, well prepared specifications should improve relationships between producer and purchaser. Each knows from the specification what is required, how it is determined, and the amount and kind of evidence necessary to make a decision. The confusion and resentment which arises because no clear-cut understanding exists as to what is acceptable or unacceptable material is eliminated.

A cogent reason for purchased material specifications is that improved quality means significant savings in manufacturing at the purchaser's plant, not only in processing but in reduced seconds and rejects. Those in the textile industry who fabricate structures from purchased yarns such as carpet manufacturers, weavers, and knitters can vouch for the very considerable added costs arising from non-standard weight, strength, moisture, twist, oil content, color, poor wind, et cetera of purchased yarns. An example of this is shown in the attached illustration, Figure 1. This is a weight or yarn size control chart for purchased yarn. Each dot on the upper chart represent the average of the tests made on one shipment. The bottom chart is a plotting of the average of all shipments received in a particular month.

It can be seen that for the first twelve months the supplier delivered yarn considerably overweight. Since he had a practical monopoly of the yarn in question at the time it could not be rejected nor

FIGURE I

WEIGHT CONTROL CHART



would any settlement for overweight be entertained. However, the supplier finally agreed to have quality control engineers from the purchaser's plant go to his mill and help organize quality controls. This occurred in November of the first year. Soon thereafter production was brought to standard weight and was kept in good control. Since the yarn was bought by the pound and used by the yard the loss to the buyer in the first year through overweight was approximately $60,000 and accordingly, savings were made at this rate in the subsequent periods of good control.

The purchasing agent's job is to obtain the required amount of raw material of the proper quality, at the right time and at the lowest possible price. The job of production is to produce the required amount of goods at the right time, of the proper quality and at the lowest possible cost. If the purchasing agent and production people have different concepts of what raw material quality should be and one dominates the other, then either way the mill can lose. Production people press for material of highest quality (and generally of highest cost) to get maximum operating efficiencies. Purchasing aims to get the lowest possible material cost (and generally lower quality). There must be a meeting of minds between Production and Purchasing resulting in a specification. This is particularly necessary in large mills, where the purchasing agent may not be in close touch with daily production problems. Here, the Quality Control Department is the liaison, the co-ordinator, whose function is to see that the thinking of both groups is crystallized in action toward quality control, i.e., the purchased material specification.

Widespread use of specifications by the purchasers of textiles will stimulate efforts toward improved quality control in the producing mill. Ultimately, the control of quality in the producing mills could reach such an effective level that the purchaser would use the producer's plant tests as evidence of quality of the finished product. His own testing activities would then be greatly reduced. This has been the case in other industries and can be so for textiles.

Process and product specifications differ from purchased material specifications in that they do not generally list lot acceptance requirements. A process specification should include:

1. List of the materials used.

2. Description of the equipment used and how it functions.

3. The formula to use for mixing the ingredients, for example, the scouring solution in washing; the fiber ratios and weights to use in blending.

4. The process requirements, that is, how the materials are combined in the processing equipment in respect to time, temperature, pressure, rate, et cetera.

5. The required standards that the process must meet and the methods of test by which conformance to these standards is judged.

Most quality control in the textile manufacturing plant is process control. Unlike the mechanical and electrical industries there is little possibility of inspecting lots in process, rejecting the bad ones, and then sorting the good from the bad. The aim is to keep the process in control, taking corrective action when tests show significant departures from standards.

Product specifications (as well as process specifications) are vital information for the cost department as well as directions for the manufacturing departments. They should be detailed enough to satisfy both groups. Product Specifications should include:

1. Identification of the grade or product, name, code number, et cetera. Date of issue should be given.

2. A brief description of the product and its use (particularly if it is made for a special use).

3. Construction - for example, type of weave, warp ends per inch, picks per inch, et cetera.

4. Materials used - the various yarns used should be listed, and references by their specification numbers, and in such a way as to clearly identify them.

5. Finishing Operations are often specified on the product specification, such as the type of back size (in carpets) or soil resistance treatment, et cetera.

6. Standards for weights of component materials are often specified in textiles as well as standards for certain properties necessary to proper functioning, for example, tensile strength, color fastness, et cetera. Methods of test for checking conformance to these standards should be listed.

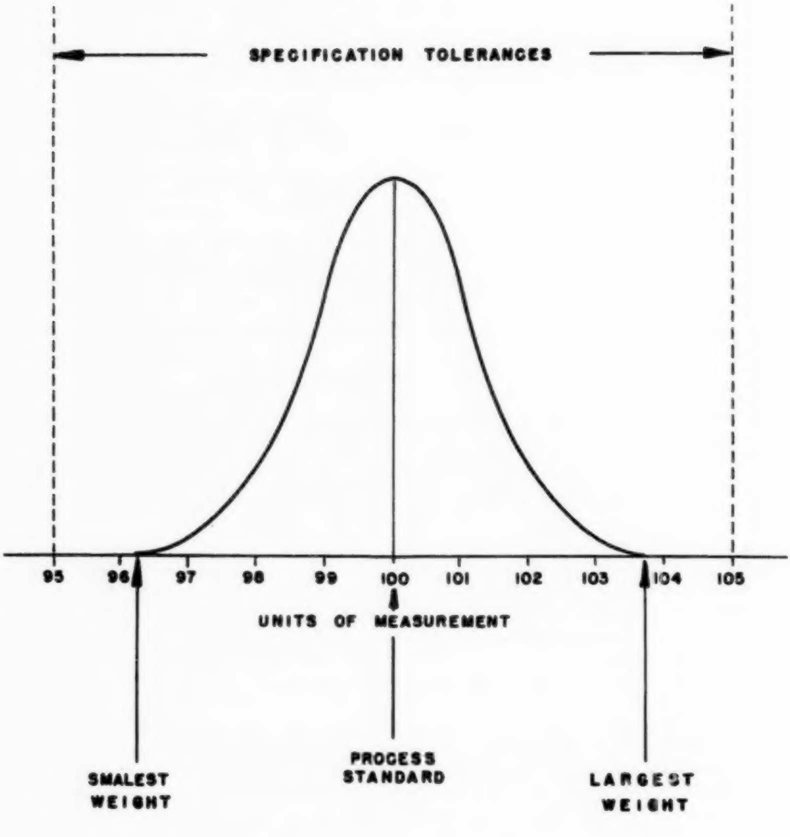
7. Packing and packaging should be described. The appearance of packages and the protection they afford are of considerable importance. A poor package can suffer greatly in transportation so that it looks sloppy on arrival, and in addition, the goods themselves can suffer damage or be creased and rumpled in appearance.

The setting of process and product standards and tolerances depends on two things: (1) the levels necessary for product performance; and (2) the capabilities of the equipment or process. These two conditions are illustrated in Figure 2. The bell-shaped frequency curve represents what the process is capable of doing when it is statistically controlled. Its horizontal spread at the base shows how much the process varies from one side of standard to the other. The dotted vertical lines are the tolerance limits which, if exceeded, cause trouble later on in the process or in the product performance. If a process has a spread greater than the tolerance limits then one of two things must be done, either the process must be changed and improved so that its variability is decreased to the point where it falls within the tolerance limits, or the tolerance limits must be increased. The action to take depends, of course, on the particular circumstances and the costs and risks involved.

FIGURE II

DISTRIBUTION OF WEIGHT DETERMINATIONS

WOVEN TEXTILE FABRIC



If such action is not taken, a situation results where a certain percentage of material is generated by the process which does not meet the tolerance limits. The natural reaction of production people is to conclude that the process has gone wrong and attempt to correct it. But this is a waste of time and a source of frustration. Unless the process is changed and improved, a percentage of substandard material will be a constant occurrence.

In textiles, it frequently occurs that tolerance limits are necessary on only one side of the standard. For example, the purchaser may specify a minimum weight, minimum strength, minimum wool content, et cetera, or that the oil content shall not exceed a certain amount. In these situations, the most economical standard for the manufacturer to set depends on the variability of his processes. If a certain minimum weight is to be met, say, in supplying the Armed Forces, and the total variability arising from spinning, weaving, et cetera amounts to ±5%, then the manufacturer's weight standard for the product must be 5% above the minimum weight given in the customer's specification or run the risk of rejction. Obviously, if a manufacturer can cut his process variability he can bring his standard closer to the specification minimum. Considerable savings can result from this. Good quality control pays off here.

The setting of process and product standards necessitates study of process capability. This is an application of statistical quality control methods, primarily control chart techniques.

Standards for purchased material, processes, and finished product, should not be the creation of one department. They should represent the best thinking of manufacturing, research and development, engineering, styling, quality control, purchasing and other interested groups. In large mills, research and development generally supplies the basic data for the preparation of a new specification. The quality control department then drafts the specification.

Drafts of specifications are reviewed by a standards committee composed of representatives of the above named departments. The chairman might be the plant superintendent, and the secretary the head of the quality control department. Each new specification or change is discussed by the standards committee and approved or modified. If approved, it is given to the general manager or the plant manager for his approval and signature.

An important responsibility of the quality control department, besides drafting specifications, is to act as custodian of the specifications, issuing them to authorized personnel, replacing obsolete specifications, and keeping a file of both past and present specifications. There are a number of points in the mechanics of operating a specification system that should be considered in any installation. They include the following:

1. Recipients of specifications. Since the information in specifications is likely to be confidential, there should be a top policy decision as to who is to receive them. Depending on the size of the mill, it may be necessary to have a written receipt for each specification.

2. Numbering, coding, identification of the specification is important. If this is not clear, costly mixups can occur where a specification close to, but not the same as, the one in question is used.

3. When a changed specification is issued, the one it replaces should be noted on it, the change that was made, and the authority for the change. All specification changes should be approved by the standards committee and the plant manager.

4. A technique should be set up for periodic review of specifications, particularly those for product, to eliminate the inactive ones. This can generally be done by consultation with the production scheduling department. All specification recipients are then notified and requested to return inactive specifications to the quality control department.

The successful use of specifications and standards rests, in the last analysis, with plant management. Good management will make clear to supervisors and workers the value of specifications and will insist that they be adhered to, and that no product be made without a specification. Improved quality of purchased material, better vendor-purchaser relations, reduced manufacturing costs, improved product quality, and tighter co-ordination of line and staff groups is obtained with the skillful use of soundly prepared specifications and standards.

REFERENCES:

1. A.S.T.M. Standards on Textile Materials, American Society for Testing Materials, Philadelphia, Pennsylvania, November, 1953.

2. Tentative Recommended Practice for Interlaboratory Testing of Textile Materials, D990-52T, A.S.T.M. Standards on Textile Materials, November, 1953.

3. Military Standard 105A Sampling Procedures and Tables for Inspection by Attributes, United States Government Printing Office, Washington, 1950.

4. Tentative Methods of Core Sampling of Wool in Packages for Determination of Percentage of Total Wool Fiber Present, D1060-53T, A.S.T.M. Standards on Textile Materials, November, 1953.

5. Dodge, Harold, and Romig, Harry F., "Sampling Inspection Tables", John Wiley & Sons, Inc., New York, 1944.

ON THE ANALYSIS OF PLANNED EXPERIMENTS

by

Milton E. Terry
Bell Telephone Laboratories, Inc., Murray Hill, N. J.

Over thirty years, two theoretical approaches to the statistical treatment of research and development problems have evolved. It is the purpose of this paper to show how both can be used together in the analysis of data.

W. A. Shewhart and others have considered the problem of analyzing process data where the number of measurements is large. The approach proposed by R. A. Fisher is to select a group of variables and a set of values of each variable, and then take measurements at selected combinations of these values in order to estimate the effect of changing each variable among its selected values, this effect being averaged over the selected values of each of the other variables. Randomization is used to average out the effect of the variables not under study.

The Shewhart method of analyzing data uses control charts wherein the data is first plotted in the pertinent recorded order in rational subgroups, and the applicable control limits found from an average "within subgroup" estimate of dispersion. A subgroup central value, and a dispersion estimate are plotted on charts together with their appropriate control limits. It is then standard to scrutinize all the charts for evidence of non-randomness and lack of control. When the data finally passes all the tests of interest, estimation is justified. Of course all datum points, and statistics not satisfying a test criterion, must be examined carefully by the research team for assignable causes. When the process yielding the data is not in control, estimation and prediction are hazardous.

Shewhart[2] points out that his recent research reveals that one may find sets of data which satisfy all simple statistical tests but display recurrent patterns which cast doubt on any hypothesis of randomness and independence. One of the most common patterns he has found occurs in the field of multiple readings with no reference point where he observes series of readings forming trend lines of varying length and magnitude of slope, with sharp breaks between segments. When the variation of these lengths and slope magnitudes is small, certain inferences can be made. When the variation is large, it is not clear what inferences should be made or with what confidence.

The analysis of a statistically designed experiment using the classical form of the analysis of variance depends on three basic assumptions of (1) additivity of

treatment effect, (2) independence, and (3) homoscedasticity. Under these assumptions it is possible to incorporate into almost all research projects a schedule of measurements on specified elements of the experiment involving the selected variables in such a way that the effects of each selected variable averaged over the combinations of selected values of the remaining variables can be measured. In addition the reality of effect from a selected variable can be tested statistically. In fact, the testing of apparent reality of effect and estimation of residual variation has been the main functions of the analysis of variance, and until recently were considered a satisfactory ending to the reduction of experimental data. Hence, some engineering and industrial research personnel have cast aside the statistical design of experiments, since they could neither satisfy all of the assumptions nor accept the classical form of the analysis of variance as satisfactory at the end of most experiments where several or all of the following questions must be answered.

Q1. Are there any assignable causes of variation present other than those introduced into the experiment deliberately?

Q2. How important are the effects of each of the selected variables?

Q3. Was the experiment well conducted?

Q4. Were there any unusual outcomes worthy of study?

Q5. How large a fluctuation can be expected in the process for manufacturing a product of which the experimental units were originally presumed representative?

Q6. What specifications can be written?

Q7. Which of the selected variables have effects demonstrated by this experiment not to be zero?

The control chart technique gives answers to these questions, but not all have the same efficiency. The analysis of variance seemed designed to answer Question 7 only, but with the aid of recent developments (components of variance, multiple comparisons, and the analysis of residuals) now offers reasonable answers to the remaining questions.

Under the assumptions of a statistically designed experiment we can always state a mathematical model. Consider the following hypothetical simple experiment. We wish to study the effect of reducing corrosion by evaporating a metal p mils in thickness on an electrical element. Ten elements at each of six thicknesses ($p_1 \cdots p_6$) are considered necessary. Only one element at a time can be coated, so the

sixty units will be processed in a random order. They are to be subjected to a controlled corrosion attack and then measured. Let t_i be the true relative effect of thickness p_i in reducing corrosion ($\sum_i t_i = 0$). Let μ be the true average corrosion effect over the experimental range, and y_{ij} the measurement of the j^{th} element with the thickness coating p_i. Then our mathematical model is

$$y_{ij} = \mu + t_i + e_{ij}; \quad i = 1, \ldots, 6; \quad j = 1, \ldots, 10$$

where e_{ij} is the residual effect and is assumed to be a random independent normal variate.

We can estimate μ by the over-all mean $\bar{X} = \frac{\underset{ji}{SS}\, y_{ij}}{60}$; and t_i by $\bar{X}_i - \bar{X}$, where $\bar{X}_i = \frac{\underset{j}{S}\, y_{ij}}{10}$. Then we define $Y_{ij} = \mu + t_i$, $(i = 1, \ldots, 6)$ to be the predicted value, and $z_{ij} = y_{ij} - Y_{ij}$ to be the <u>residual</u> of the measurement (ij). It follows that $\sigma^2 = \sigma^2_{z_{ij}} = Sz^2_{ij}/54$.

We simulated this experiment by assigning constants to the μ and t_i, and values to the ε_{ij} from a table of random numbers to yield y'_{ij}. Then the set y'_{ij} were placed in a random order. In two simulations, with respect to the ordered y'_{ij}, a linear trend and an abrupt shift in level were superposed respectively on the y'_{ij} to yield two sets of data y_{ij} of known behavior (see Figures 1 and 3). Standard analyses were run. The estimate of relative mean effects were not very biased, but the estimates of the residual variation were so bad that no conclusions about equality of effects could be drawn. Then the z_{ij} were calculated for each simulation and plotted against order (see Figures 2 and 4). When the data of Figure 2 was corrected for the fitted trend line, the new estimates of the known parameters were excellent. The use of Figure 4 gives an excellent estimate of the shift in level and again correctly adjusted the estimates from Figure 2.

When the set of residuals, z_{ij}, constitute a time sequence, they can be plotted as such. In many engineering experiments only one fabricating or measuring device is available, and hence one or more time sequences are imposed on the experiment. In general the statistical design will average out the time effect in the estimates $\hat{t}_i$ by randomizing the order of fabrication or measurement of the experimental units.

In a real sense, the set of residuals plotted against time together with control limits, $\pm K\sigma_{residual}$, are a control chart. Hence we are tempted to use the usual chart techniques. Since there are constraints imposed by the model, the significance levels are no longer identical with the tabular values. But when the control limits are used as action limits, satisfactory results should ensue.

Anscombe and Tukey(1) have proposed plotting the set of residuals (z_{ij}) against its associated predicted value Y_{ij}, when the experiment contains at least a double classification. Here "non-additivity is shown by a curved regression. Non-constancy of variance is shown by a wedge shape."(1)

In general plotting residuals both against their predicted values, and against serial order(s) enables the experimenter to examine that portion of his measurements which is not attributable to the suspect variables. He will have visual evidence as to the vexations from many sorts of non-additivity of effect, non-constancy of variance, linear trends, cycles, and wild shots which may be embedded in his experiment. Hence the analyst-experimenter can take the necessary action to ensure that the final accepted readings in the proper units satisfy the assumptions on which valid predictions and estimates will be made. This form of analysis, used in conjunction with the analysis of variance, enables the user of a statistically designed experiment to focus the same type of scrutiny on his data that the control engineer can give to process data.

References

(1) Anscombe, F. J. and Tukey, J. W.: "The Criticism of Transformation," 1954, unpublished manuscript.

(2) Shewhart, W. A.: Personal communication.

TESTING ONE-QUARTER MILLION TRANSISTORS

George R. Scheel and William H. Greenbaum
Sonotone Corporation

A manufacturer of electronic hearing aids is faced with the problem of maintaining high quality in its product to a degree much greater than that experienced by manufacturers of many other consumer electronic devices. The average hearing aid operates for 5000 hours a year under conditions of high humidity. The unconditional guarantee for one year on a Sonotone hearing aid, that permits the user to have his hearing aid replaced at any Sonotone office in the country for any reason, in effect means that Sonotone performs all the maintenance for one year on every hearing aid that it produces. It is obvious, therefore, that the "stay-out" ability of the product is of great economic interest.

Hearing aid production in the United States is now 100% transistorized. As a leader in the development of transistor specifications and test equipment for audio use, Sonotone, in cooperation with transistor manufacturers has developed a program of testing that includes feedback paths for data to the vendor that is unusual in industry. To date, Sonotone has received at Incoming Inspection over 250,000 transistors purchased to its own specifications. These transistors have been tested and data correlated both for our own information and for the information of the appropriate vendor. These data include incoming inspection, analysis of in-plant failures, and analysis of returns from the field. In addition, life tests are run under temperature, humidity, and power cycling conditions. Variables data are recorded and plotted, and copies of the graphs are forwarded to the individual vendors.

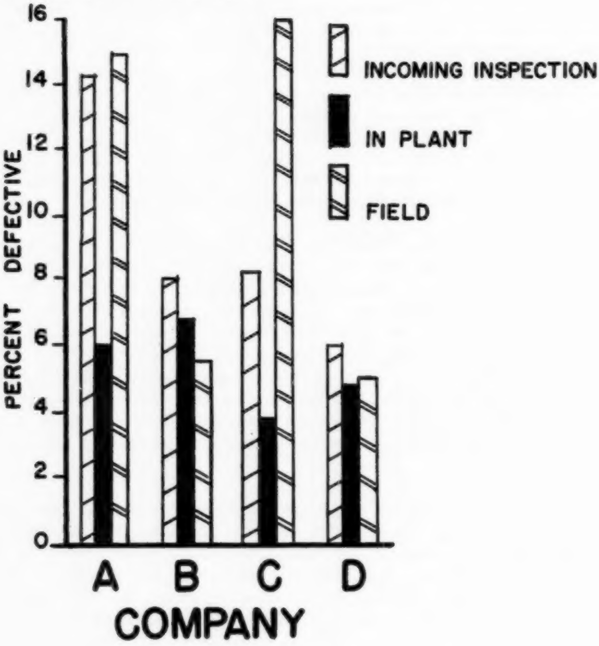
As transistors are received they are subjected to a sequence of tests designed to eliminate the most common failures first. Before testing, a slow, overnight, temperature stabilizing cycle is applied to eliminate mechanical failures due to temperature variation and to present all transistors for test with identical, immediately preceding temperature histories. The tests include (in the following order): a test for the saturation current at cut-off (Ico); a test for noise which includes a one hour noise drift test; and tests for gain under the several voltages at which the hearing aid operates. Transistors that meet specifications are color coded in gain groups for selective assembly in hearing aid production.

In production, electrical tests are performed on the wired chassis and later a complete, final test is performed on the assembled hearing aids. Hearing aids out of specification are transferred to analyzers, and defects recorded by them are corrected by repairmen. Transistors found inoperative in this sequence are returned to the same transistor test equipment used for incoming inspection and retested. Hearing aids returned from the field are tested, analyzed, and repaired in the same sequence, and all transistors removed from these instruments are also retested on the same equipment used for incoming inspection. Attributes data for each characteristic are recorded on all transistors and supplied (with the rejected transistors) to the individual vendors. These rejects are divided into three classes; those found at Incoming Inspection, Inplant, and Field Failures.

In addition, variables data are collected on each specified characteristic on a sample of approximately 100 transistors out of every ship-

ment. These data are compiled in the form of a monthly Incoming Inspection Report on Transistors and each vendor receives copies of the histograms describing his product.

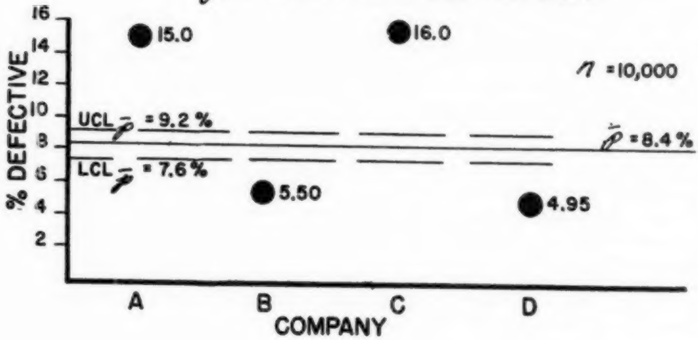
With data available on over one-quarter million transistors, we are able to observe the different behavior between transistors and vacuum tubes, and also the nature of the long term field problem of this new device. It is known by everyone who has read about the transistor that the problem of burned out filaments and microphonic tubes is non-existent. In addition, most readers would be surprised to find that the gain of the transistor is no longer a major cause of rejects at any point during the life of the hearing aid. To date, major problems are stability of saturation current (Ico), control of noise, and elimination of dead and intermittent transistors which are caused mainly by poor mechanical design or defective workmanship. To bring these facts to the attention of the reader, data have been collected for presentation based on the product of four of the more than ten vendors from whom we have purchased transistors. A minimum of 30,000 transistors have been purchased from each of the vendors discussed herein.



(FIG. 1)

In Fig. 1 a comparison of four transistor manufacturers has been made on the basis of the percentage defective at Incoming Inspection, the percentage defective during the in-plant period, and the percentage of defective transistors that have been returned from the field in one year. It will be noted that Company A had the poorest incoming inspection level, but Company C, with approximately half as many incoming re-

"p" CHART OF FIELD RETURNS



3σ Control Limits For $\bar{p} = \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

$$= .084 \pm 3\sqrt{\frac{.084 \times .916}{10{,}000}} = .084 \pm .0083 = 8.4 \pm .8\%$$

"t" Test For Significance Of Difference In "p" Of Companies B And D (See Page 384 Ref. 1)

$\bar{p}' = \frac{p_B + p_D}{2}$, Since $n_B = n_D$, $\bar{p}' = \frac{.0550 + .0495}{2} = .0523$

$$t = \frac{p_B - p_D}{\sqrt{\bar{p}'(1-\bar{p}')\left(\frac{n_B + n_D}{n_B n_D}\right)}} = \text{But } n_B = n_D, \therefore t = \frac{p_B - p_D}{\sqrt{\bar{p}'(1-\bar{p}')\frac{2}{n}}}$$

$$t = \frac{.0550 - .0495}{\sqrt{.0523(1-.0523)\frac{2}{10{,}000}}} = 1.73$$

t	PROBABILITY
1.64	.90
1.96	.95

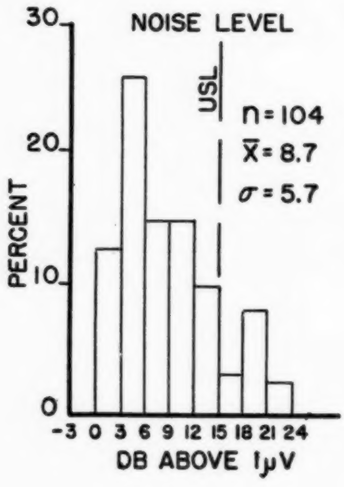
∴ BY INTERPOLATION,
FOR t = 1.73,
PROBABILITY = .92

FIG. 2

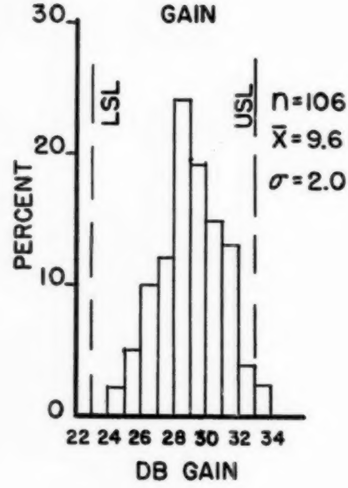
jections and the lowest level of in-plant failures, had the most field failures. It is apparent that Company B and Company D had a lower level of failures than Company A and Company C. In Fig. 2 we have plotted a "p" chart of the field returns for Company A, B, C, and D. It will be noted that Company A and C are significantly different from Company B and D, and that all companies are outside control limits for $\bar{p}$, the combined performance of all transistors.

Let us focus our attention on the two best products, those of Company B and D. Are they the same or are they statistically different? By application of Student's "t" test for significance of differences between proportions, we conclude that if these two samples were drawn from a single population there would be eight times in a hundred that this difference would occur by chance. As a result of this test one could not conclude that these two products are statistically different. (See Fig.2 for application of "t" test). These data as shown in Fig. 1 and Fig. 2 can certainly assist one in deciding which vendors make the better products, but do not indicate which vendor makes the best product. A further look into the detailed analysis or rejection by causes will be helpful.

When a shipment of transistors is received from the vendor, variables data are collected on a sample of approximately 100 units. Fig. 3A shows a typical plot of the "noise distribution" in a sample, and Fig. 3B shows a typical plot of the "gain distribution" in a sample. In addition to the actual attributes data, these data are made available to the vendor.



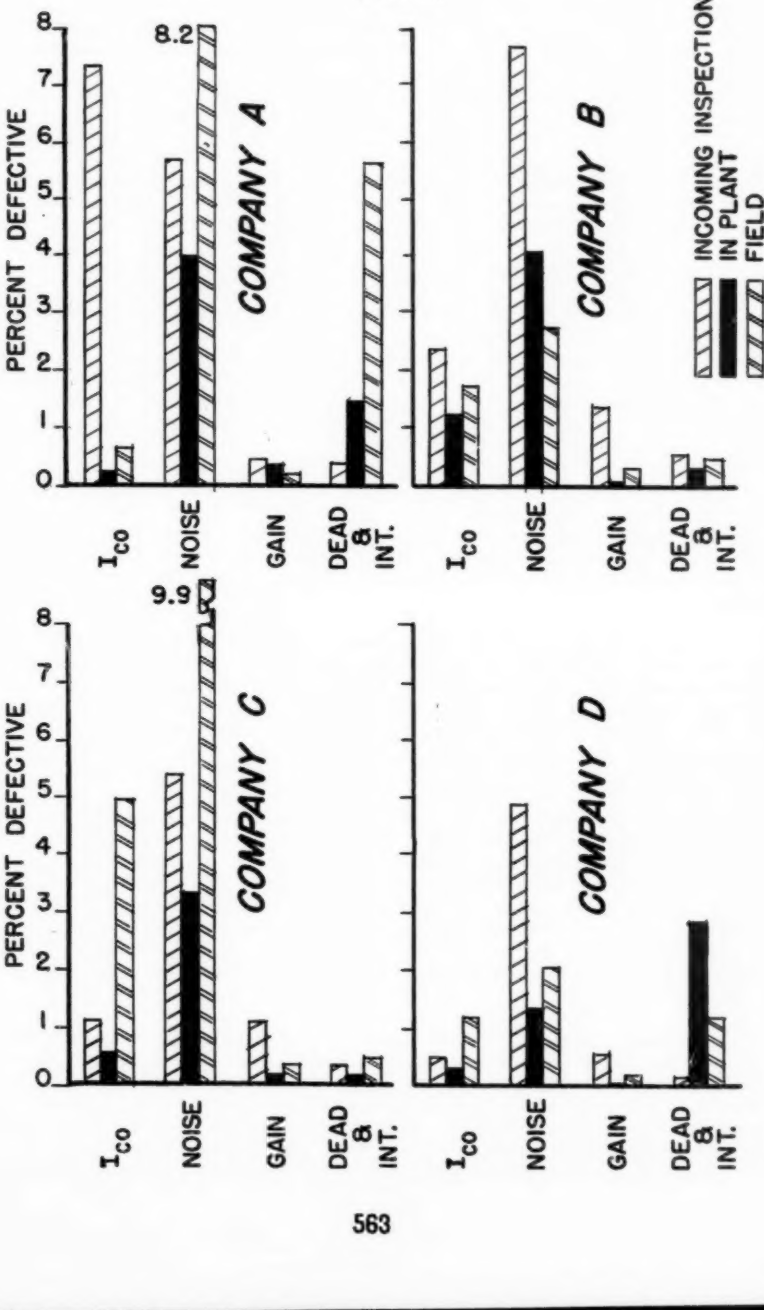
(FIG. 3A)



(FIG. 3B)

It will be noted during the discussion of Fig. 4 that these particular types of distributions (Fig. 3) determine the relative importance of the different tests. In Fig. 4 the defective transistors of each company have been separated under four major causes of rejection and have been plotted.

(FIG. 4)



COMPANY A
COMPANY B
COMPANY C
COMPANY D
PERCENT DEFECTIVE
0
1
2
3
4
5
6
7
8
8.2
9.9
I_{co}
NOISE
GAIN
DEAD & INT.
INCOMING INSPECTION
IN PLANT
FIELD

Company A has been chosen to illustrate a pattern of a product that was electrically good, but mechanically bad. The high incidence of rejection for Ico (an electrical characteristic) at Incoming Inspection, followed by few future failures for Ico, in-plant or in the field, indicates a difference in correlation of test equipment or that the vendor is testing to a higher limit. The increasing amount of defectives for dead and intermittent transistors (normally a mechanical failure) from in-plant to field, along with the large number of field returns for noise (normally an electrical failure), correlated to the findings that the transistors were actually failing because of poor mechanical structure. A condition existed where the internal wires were poorly soldered in varying degrees resulting in either noise, intermittency, or open circuit. Normal transistor experience shows that high Ico and high noise go hand-in-hand. In the case of Company A this was not so because the mechanical problem far overshadowed the intrinsic transistor behavior.

The data of Company B in Fig. 4 exhibits a pattern of a good product that maintained its quality level in the field. The high rejection rate at Incoming Inspection for noise can be related to the histograms of incoming noise, such as shown in Fig. 3A, where the product as actually manufactured was not quite the product required by specification. Inherent to the problem of measuring noise, correlation is difficult, and at best can be maintained to 1 or 2 decibels (db) between the vendor and user. These two factors resulted in the high initial rejection for noise of Company B.

The production of Company C showed a satisfactory quality level at Incoming and In-plant Inspection, but the rejection for high Ico and noise from the field was so excessive that Company C had the highest overall reject rate. Note that unlike Company A, the number of dead and intermittent transistors has not increased along with the increase in noise.

Company D exhibited the best overall quality level but still suffered from some difficulty in noise test correlation. Company D's product could almost have been termed exceedingly good, except that the incidence of dead and intermittent transistors exceeded that of either Company B or C. When the information obtained from in-plant rejections was presented to Company D, they immediately instituted an intensive campaign to improve the quality of assembly and eliminate the bad connections.

It will be noted that all four companies in Fig. 4 show rejections for gain that never exceed 1.5%. Company B and C had over 1% rejection at Incoming Inspection (which was caused by a test equipment correlation problem), however, Company B and C had less than .3% rejection from the field for low gain. This information points up one of the gratifying characteristics of the transistor, in that the gain of this device will not deteriorate in the manner of the vacuum tube. Fig. 3B shows a typical gain distribution and indicates the reason for small rejections for gain.

The data presented in Fig. 1 and Fig. 4 include field experience on these transistors for approximately one year. As the pattern began to evolve, the information was conveyed to the appropriate vendor and discussions were held to analyze the cause of failure and the steps to be taken to improve the product. In most cases cooperation was excellent; occasionally pressure had to be applied. By such major pressures as having one company stop shipment for several months, completely refusing

to buy products of another company, and having a third company change its mechanical design and packaging, a great improvement in the product has been taking place.

The transistor is as yet far from perfect, but the latest data and the latest field experience indicate that transistors are now superior to the vacuum tubes used in hearing aids.

References

(1) Juran, J. M., Quality Control Handbook, New York: McGraw-Hill Book Co., Inc., 1951, Page 384.

QUALITY CONTROL APPLIED TO PLATING

Guy J. Campbell
Ternstedt Division
General Motors Corporation

The scope of this paper is confined to the specific application of chart control to the electroplating process. This usage is one of the many adaptations of Statistical Quality Control techniques to be found in our organization.

The Ternstedt Division of the General Motors Corporation specializes in the manufacture of automotive trim and hardware; and, among the many items produced by the Detroit plant are various plated parts. Because the rejects for the copper-nickel-chrome plated articles were relatively high, considering the cost involved in the salvage operations, it was one of the first phases of manufacturing to receive the attention of the Statistical Quality Control Department when it was formed.

First, Percentage-Defective Charts were installed at three stages in the manufacturing cycle to portray the conditions as they actually existed. The charts covered the fabrication and bare-metal-finishing rejects, the copper plate and buff rejects, and the defectives found after nickel-chrome plate. A breakdown of the rejects, by type of defect, was made an integral part of the chart to aid in directing corrective action. For simplification, the rejects after the plating operations were grouped into three main classifications: (a) fabrication defects which had not been detected during previous inspections, (b) handling defects such as scratches, nicks, etc., and, (c) defects caused by the plating process. To avoid straying from the subject, the problems encountered in securing a reduction of rejects in the first two groups will not be discussed here.

In order to gain a closer control of the quality of the plate, check points were established for the floor inspectors at each plating conveyor. However, a resulting increase in down time, necessary to effect corrective measures, posed another serious problem.

To solve it, consideration was given to the possibility of using charts as an aid in controlling the plating solutions. It was known that the balance and concentration of the solutions were prime factors in determining the quality of the plate. For example, weak cleaners would not remove all of the dirt and buffing compounds from the part; also, the acids and the metal solutions had to be maintained within certain limits in order to obtain the proper thickness and adherence properties.

Daily tests by the conveyor operators of the acids and cleaners, and weekly analyses of the other solutions by the chemical laboratory were already being performed and the results duly recorded and filed. Preliminary charting of this data indicated that the recommended specifications of the laboratory were often ignored. Such was the case with the cleaners, where it was noted that the concentration was consistently increased day by day, although not shown necessary by the analysis. Another practice was to add material to the various tanks as soon as defective plate from an unknown cause was encountered; as a consequence, the specification limits were many times violated. Of particular

interest was the fact that the conveyor operators would continue to make daily additions of chemicals on blanket instructions, disregarding the results of the analyses.

It was felt that conditions such as these could best be relieved by acquainting everyone concerned with the results of the analyses and the proper specifications. To accomplish this end, charts were established for every solution in a plating conveyor. The specifications were indicated by red lines to serve as upper and lower control limits. Because an individual solution is (for all practical purposes) homogeneous, only one sample is necessary for any one check; therefore, the specification limits become the theorectical control limits. As with theoretical limits on other types of charts, it was necessary to change them (and the specifications) as the process warranted it. By displaying the charts near the conveyors, the conveyor operators, together with others who were concerned with the operation, became better informed as to the desired strength and the results of the analyses. Through insistent questioning of every addition not indicated as necessary by the charts, the solutions were eventually brought within control. This has not only resulted in improved plating, but also, a significant saving in material. The following examples illustrate the progress that has been accomplished by this program in our plant.

Figure 1

The top chart of Figure 1 shows the "before" condition of cleaners which was previously mentioned. The lower chart reveals that the practice of steadily increasing concentration has been discontinued. Now, only sufficient material is added to the solution to stay within the specification. The difference in the amount of cleaner used is approximately 30% for the two months represented by the charts.

Figure 2

Figure 2 is another example of waste. In this case, it was due to blanket additions of 12 gallons of muriatic acid each day, even when the concentration was above the upper specification limit. The lower chart is for the succeeding month when the chart was the controlling agent for the additions. Note that no additions were necessary for 15 operating days.

Figure 3

The top chart of Figure 3 features a state that was also noted. Although daily additions were being made, they were inadequate, for the most part, to achieve the desired strength. By the present method of adding the necessary amounts to obtain the correct concentration and then only making sufficient addition to keep it within specification, a material saving is again accomplished, together with improved plating.

Figure 4

As with other operations, the specifications for electroplating solutions sometimes require revision to improve the process. This is exemplified by the charts in Figure 4, where the limits for sulphuric acid were changed to eliminate a latent peeling plate condition which was being experienced with the original specification.

Another application that has proven helpful is the charting of defective plating racks. The information is furnished by the floor inspector who, daily, selects a random sample of twenty-five to fifty racks from the monorail leading to the plating conveyor. He inspects them for defects such as "treed" plate, broken insulation, bent or broken retention clips and other conditions which contribute to poor plate. All of the defective racks which are found are set aside for repair and the percentage of defectives is posted on the chart. The public display of this chart has not only acted as an incentive to the Rack Repair Department, but also, was indirectly responsible for securing a better coating (insulation) for the racks. By improving the condition of the racks, the number of misplated parts has been substantially reduced and waste of plating metal due to "treeing" has been decreased.

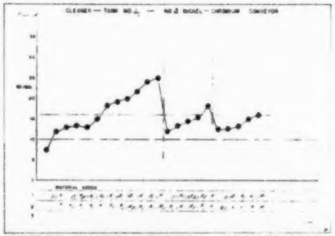
In conclusion, it should be emphasized that this method of chart control will not solve all of the plating problems; the process, itself, is too complex and is, to a degree, influenced by outside factors such as, atmospheric temperature and pressure. However, it has been successful in promoting control of a main component of the operation, which is the chemical concentration, to improve quality and save material.

This is another instant where we have profited by applying two basic principles of Statistical Quality Control: (1) chart the information and exhibit it publicly where all those concerned can see it, and (2) although variation exists, it can be controlled.

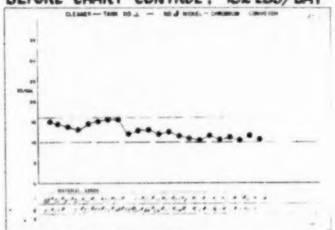
TERNSTEDT QUALITY CONTROL

DIRECT CLEANER ~ NICKEL-CHROME CONVEYOR

TOP CHART SHOWS OUT OF CONTROL CONDITIONS AND ALSO LARGE ADDITIONS OF MATERIAL WITHOUT APPRECIABLE CORRECTION. WITH QUALITY CONTROL PROGRAM, SOLUTION WAS KEPT WITHIN SPECIFICATIONS BY ECONOMIC AND SYSTEMATIC ADDITIONS OF NEEDED MATERIAL.



BEFORE CHART CONTROL ; 182 LBS/DAY



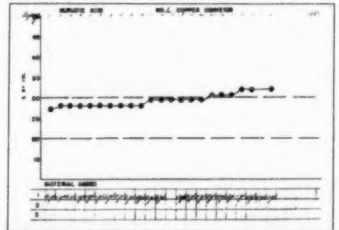
AFTER CHART CONTROL ; 125 LBS / DAY

Figure 1

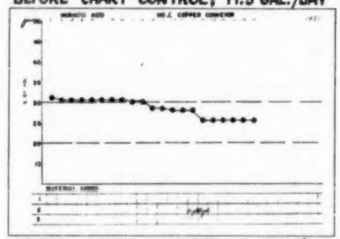
TERNSTEDT QUALITY CONTROL

MURIATIC ACID ~ COPPER CONVEYOR

PRIOR TO CHART CONTROL, DAILY ADDITIONS OF ACID WERE MADE WITHOUT REGARD TO REQUIREMENTS. INSTALLATION OF A CHART RESULTED IN THE PROPER CONTROL OF THE SOLUTION WITH A MINIMUM OF ADDITIONS AND CONSEQUENT MATERIAL SAVINGS.



BEFORE CHART CONTROL; 11.5 GAL./DAY



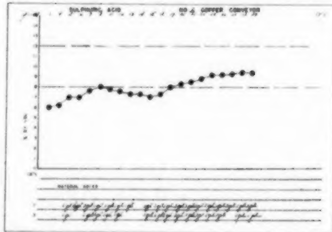
AFTER CHART CONTROL; 1.1 GAL./DAY

Figure 2

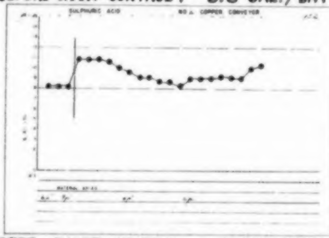
TERNSTEDT QUALITY CONTROL

SULPHURIC ACID ~ COPPER CONVEYOR

PRIOR TO CHART CONTROL ADDITIONS WERE HABITUAL AND INADEQUATE TO BRING SOLUTIONS UP TO DESIRED STRENGTH. WITH THE ADVENT OF QUALITY CONTROL METHODS THE SOLUTION WAS MAINTAINED WITHIN SPECIFICATIONS WITH A FEW TIMELY ADDITIONS.



BEFORE CHART CONTROL ; 3.3 GAL./DAY ~



AFTER CHART CONTROL ; 1.3 GAL./DAY ~

Figure 3

TERNSTEDT QUALITY CONTROL

ACID COPPER PLATING ~ COPPER CONVEYOR

UPPER CHART ILLUSTRATES QUICK COMPLIANCE TO A SPECIFICATION REVISION, WHICH WAS NECESSITATED BY A LATENT PEELING PLATE CONDITION. LOWER CHART, WITH THE REVISED SPECIFICATIONS, SHOWS A RECENT DAILY ANALYSIS OF SOLUTION WITH CONTROL MAINTAINED BY A MINIMUM OF ADDITIONS.

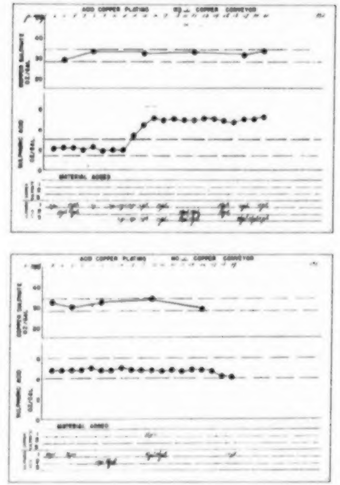


Figure 4

Statistical Sampling Methods Applied to Auditing and Accounting

John Neter
Syracuse University

Introduction

Statistical sampling methods can be of great value in auditing and general accounting applications. The use of statistical sampling methods in these fields is a relatively recent development. It is the purpose of this paper to indicate some of the areas in auditing and accounting where the application of statistical sampling techniques appears to be fruitful, to point out some of the problems which arise when such applications are made, and to cite some case histories where the application of statistical sampling techniques in auditing and accounting has proven to be successful.

First, the use of statistical methods in auditing will be considered. The application of statistical sampling methods in order to obtain accounting information efficiently will be taken up next, and finally the use of statistical techniques to control clerical accuracy and other processes will be discussed.

Auditing Applications

Nature of the problem - The purpose of the usual type of audit is to determine whether the balance sheet and statements of income and surplus present fairly the financial position of the company as of a given date, and the results of its operations for the period then ended, in conformity with generally accepted accounting principles applied on a basis consistent with that of the preceding year. An auditor, in the course of his examination, employs sampling in many instances because 100 percent examinations would often be prohibitively expensive and much too time-consuming. An auditor, for example, may sample vouchers, accounts receivable, inventory, postings, canceled checks, sales invoices, etc.

Whenever an auditor uses samples, he is faced by important sampling problems, such as the determination of the necessary sample sizes and the interpretation of the sample results. These sampling problems have caused great concern to auditors, especially in view of the auditors' professional responsibilities to clients and third parties. Up to now, however, auditors have generally relied upon judgment samples, which do not permit objective answers to these sampling problems. Statistical sampling methods may therefore prove to be quite useful to auditors in the handling of their sampling problems.

Use of acceptance sampling plans -- In order to be able to apply statistical sampling methods, the auditor must develop quantitative measures of the audit results in which he is interested. This, undoubtedly, is one of the major problems in the application of statistical sampling techniques to auditing because auditors presently do not generally use such quantitative formulations of the audit results. These quantitative measures of the audit results must be developed according to the purposes of the particular audit step within the framework of the over-all audit purposes. One possible type of quantitative measure which might be useful to auditors is the proportion of items which are incorrect, in error, require investigation, or possess some other speci-

fied characteristic or characteristics.

For instance, an auditor may be interested in the proportion of a year's purchase invoices which are incorrect. If the auditor wishes to use this quantitative measure, he must carefully define when a purchase invoice is to be considered correct and when it is to be considered incorrect. Problems will arise in the formulation of this definition since it is essential that the error definition be meaningful to the auditor. Suppose that a purchase invoice is to be regarded as incorrect if there is an error in the amount; otherwise it is to be considered correct. Note that a purchase invoice is to be considered incorrect with this definition whether the dollar error is $1 or $100. Such a formulation is a meaningful one when interest centers chiefly upon the extent to which the various factors inherent in the accounting process lead to errors. In that case, the actual magnitude of the dollar error may not be directly significant if different amounts of errors are associated with the same cause. For instance, the significance of a transposition error would probably be the same, whether the error involved is small or large, if the same basic causes were responsible for the error in either case.

An auditor may use acceptance sampling plans in conjunction with the quantitative measure "percent of purchase invoices which are incorrect" if he wishes to determine whether or not the quality of the year's purchase invoices is satisfactory. In that case, he must specify a satisfactory and an unsatisfactory level of this percentage. Here again, the auditor will face problems because he must specify these percentages so that they will be meaningful to the purposes of that audit step within the general framework of the over-all audit purposes. Suppose that the auditor decides that an error rate of 1 percent or less is satisfactory, and that an error rate of 4 percent or more is unsatisfactory.

Whenever a sample is used to make a decision, such as whether or not the quality of the year's purchase invoices is satisfactory, risks exist that the sample will lead to an incorrect conclusion. For instance, one may conclude on the basis of the sample that the quality of the invoices is satisfactory when actually it is unsatisfactory. Again, one may conclude on the basis of the sample that the quality of the invoices is unsatisfactory when actually it is satisfactory. These sampling risks of incorrect decisions cannot be avoided unless a 100 percent examination is conducted. With judgment samples, the sampling risks exist but cannot be evaluated. With statistical, or probability, samples, on the other hand, these risks can be evaluated and, indeed, can even be specified in advance.

Suppose that the auditor specifies that he cannot assume more than a 5 percent risk of concluding that the purchase invoices are unsatisfactory when actually they are satisfactory. Suppose further that the auditor specifies that he can accept at most a 1 percent risk of concluding that the purchase invoices are of satisfactory quality when actually they are of unsatisfactory quality. A statistical sampling plan can then be determined which will meet these specifications as to the maximum risks of being led to incorrect decisions. For the above case, the appropriate sampling plan would be, assuming that the number of purchase invoices for the year is large: Select at random 398 purchase invoices and examine them to determine their correctness. If 7 or less invoices in the sample are incorrect, conclude that the quality of the year's invoices is satisfactory; if 8 or more invoices in the

sample are incorrect, conclude that the quality of the year's invoices is unsatisfactory. The auditor, by following this sampling plan, will then be assured that he is incurring no more than the previously specified risks of making incorrect decisions - namely, a maximum risk of 1 percent of concluding that the purchase invoices are satisfactory when actually the error rate is 4 percent or more, and a maximum risk of 5 percent of concluding that the purchase invoices are unsatisfactory when actually the error rate is 1 percent or less.

It should be noted that a random selection of the sample of purchase invoices is essential in order that this sampling plan provide the specified assurances against incorrect decisions. Practical problems in the selection of random samples in auditing may arise, but time does not permit a discussion of them here. Suffice it to say that random selection of samples has been found feasible in many other areas of application.

Use of statistical estimation procedures - If the auditor is principally interested in the magnitude of the dollar errors or "differences" in a set of accounting records, a different quantitative formulation of the audit results would be needed because the above approach does not explicitly recognize the magnitude of the dollar errors involved. Let us consider a quantitative measure which does explicitly recognize the magnitude of the dollar errors. Suppose that an auditor wishes to verify the accuracy of the dollar value of the inventory on hand, which is recorded on the books at, say, \$5.7 million. The auditor therefore selects a sample of inventory items, determines the quantity on hand for each of these selected items, prices them, and from this sample information wishes to estimate the audited value of the inventory which he would have obtained if he had made a 100 percent examination of the inventory.

Here, then, is a measure which can serve as a basis for determining the extent of the dollars errors in the book valuation. Unless the sample is a probability sample, however, the auditor will not be able to use it in order to decide whether the sample result is useful for evaluating the accuracy of the book figure. To see why a probability sample is needed for this purpose, suppose that the sample estimate of the total audited inventory value is \$5.2 million. Since this is only a sample estimate, it will differ in all likelihood from the audited value which would have been obtained with a 100 percent examination. Suppose that it were known that the sample estimate does not differ by more than ± \$.75 million from the audited value which would have been obtained with a 100 percent examination. It could then be concluded that the audited value of the inventory is somewhere between \$5.2 ±.75 million, or between \$4.45 million and \$5.95 million. The auditor might well feel in that case that the sample estimate is not precise enough to be of much help to him in evaluating the accuracy of the book figure. On the other hand, suppose that it were known that the sample estimate does not differ from the audited value which would have been obtained with a 100 percent examination by more than ± \$.1 million. In that case, it could be concluded that the audited value of the inventory is somewhere between \$5.1 million and \$5.3 million, and the auditor might then consider this estimate to be precise enough to help him in evaluating the accuracy of the book value of the inventory.

With judgment samples, the error range or precision of the sample estimate cannot be evaluated from the sample. Thus, the auditor would

not be in a position to decide whether or not the precision of the estimate is high enough to enable him to evaluate reasonably the accuracy of the book value of the inventory. With probability samples, on the other hand, the precision of the sample estimate can be evaluated. More than that, the auditor can in advance specify the precision of the estimate which he requires; for instance, he might declare that he needs an estimate of the audited value of the inventory with an error range of no more than ± \$.2 million. With such a specification, a statistical sampling plan can often be developed then which will provide the desired estimate with approximately the specified precision. The use of information from past experience can be of great help in designing a sampling plan which will provide the required precision at as small a cost as possible. Many problems of a technical nature exist in planning the sample - for instance, in choosing the best method of estimation, the most appropriate method of sample selection, and so on. While these problems are too extensive to be discussed here, it should be pointed out that appropriate statistical sampling methods to estimate given characteristics in an efficient manner have been developed for many different areas. Therefore, there is good reason to expect that these methods, or others to be developed especially for the accounting and auditing area, should also be helpful in obtaining information with specified precision economically for accounting and auditing uses.

One other matter in connection with evaluating the precision of a sample estimate should be discussed briefly. Conclusions based upon sample results can never be certain. Thus, one cannot be certain that the error range for a sample estimate will actually include the value which would have been obtained with a 100 percent examination. A degree of assurance only can be attached to the statement that, say, the audited value of the inventory is somewhere between \$5.1 million and \$5.3 million. Suppose that the degree of assurance for this statement is .99; this means that a procedure has been followed which leads to correct statements 99 percent of the time. Thus, confidence can therefore be had that the above statement is a correct one. The degree of assurance can be specified by the auditor. In general, the higher the degree of assurance which he desires for an estimate with a specified precision, the larger will be the required sample size with a given sampling procedure.

Some applications of relevance - Some actual applications will now be described in which use of the above statistical sampling techniques was made either for auditing purposes or for other purposes quite parallel to those of auditing.

The Philips Group of Electrical Companies in Great Britain is using acceptance sampling plans in the internal audit of purchase invoices of all types, of petty cash and similar transaction, and of stock lists. (Ref. 1). The verification of purchase invoices, for example, is done monthly in most of the companies. Since the invoices are numbered serially, a random sample can be selected rather easily by means of tables of random numbers. An invoice is considered incorrect if any one of a number of different possible errors in handling it was committed. The range of these possible errors illustrates the intensive type of audit which is carried on for each invoice in the sample. Among the error possibilities are: 1) allocation made to the wrong nominal ledger account; 2) allocation made to the wrong costing code or, where applicable, to the wrong product group; 3) invoice not chcked with copy of purchase order, or accepted at price different from that on order with-

out authority; 4) invoice not checked with goods received note, or not checked correctly; 5) incorrect standard price calculated for the invoice; 6) invoice incorrect arithmetically by an amount which should have been adjusted after contact with supplier; 7) passing of two invoices for the same charge; 8) credit not claimed when it should have been claimed; 9) various errors made on items such as packing cases, transportation, custom duty, etc.

An error rate of 0.5 percent or less in the purchase invoices is considered satisfactory, while an error rate of 5 percent or more is considered unsatisfactory. The maximum risk of concluding that the quality of the purchase invoices is satisfactory when actually it is unsatisfactory was set at 10 percent, while the maximum risk of concluding that the quality of the purchase invoices is unsatisfactory when it is really satisfactory was set at 5 percent. From these requirements, the appropriate acceptance sampling plan was determined.

If the sample indicates that the quality of the purchase invoices is unsatisfactory, an attempt is made - either by additional verification or by other auditing procedures - to isolate the errors that caused the rejection and to study them as much as possible. The results of this study are then reported to the accountant in charge of the work, and serve as a basis for remedial action in the accounting department. The use of the results of the intensive audit in this manner has been an important factor in improving the quality of the accounting work in the Philips Companies.

Similar intensive sample audit procedures are employed for verifying petty cash transactions and for checking stock lists. A. C. Smith, formerly the Internal Auditor of the Companies, believes strongly that an intensive auditing of a small section of the work provides a better insight into the real state of the administration than an extensive examination of entries, with only superficial attention given to their significance. (Ref. 1). He has found that small random samples are well suited for such intensive audit examinations when combined with adequate auditing techniques. While the introduction of statistical acceptance sampling plans at the Philips Companies, together with the intensive examinations, has not led to any savings in the time taken on the audit, there has been a considerable improvement in the quality of the work. (Ref. 1).

The same satisfactory results with intensive audit examinations based upon small random samples were obtained in an audit of a county government. (Ref. 2). Several areas of the accounts were examined by random sampling procedures, including warrants payable, vouchers payable, and payrolls. The audit of payrolls, for instance, included checking of the authorization of salaries by civil service and the certification by department heads, checking of salaries with civil service personnel files, and checking of various information appearing on the payrolls with the warrants. A transaction, thus, was checked through all the papers relating to it in addition to tracing it through the accounts.

While a common auditing practice is to select, say, a week's or a month's transactions, a random sample of the entire year's transactions will usually lead to the selection of transactions scattered throughout the year. This occurred in the audit of the county government. It was found that the random samples were so scattered that they led auditors

to open files which otherwise would not have been touched. This helped to impress the county employees as to the thoroughness of the examination. The random sampling, together with the intensive audit procedures, pleased the auditors because it encouraged more care at each step of the examination and because the sample covered more areas of the accounts. (Ref. 2, p. 474).

Another use of acceptance sampling plans by auditors has been made in connection with verifying the accuracy of agings of accounts receivable. (Ref. 3). The particular concern studied was a large metropolitan department store carrying about 100,000 accounts receivable from customers. It had been the practice of the store to select about 15,000 of the regular accounts for detailed aging. The public accountant then selected about 1,500 of these accounts and checked the accuracy of the agings. The auditors decided to try statistical acceptance sampling plans to determine whether or not the store's agings were sufficiently accurate. An error rate of 3 percent or less in the client's agings was considered satisfactory, while an error rate of 8 percent or more was considered unsatisfactory. The maximum risk of accepting unsatisfactory work was set at 5 percent, while the maximum risk of rejecting acceptable work was set at 10 percent. (Ref. 3, p. 297). A sequential acceptance sampling plan was then determined which embodied these requirements as to protection against incorrect decisions. This sampling plan required, on the average, a sample of only about 126 accounts before a decision as to the accuracy of the agings can be reached, compared with the sample of about 1,500 accounts which had been selected previously.

In this same instance, the auditor also investigated the size of the sample which the client selects in order to obtain an estimate of the age distribution of all accounts receivable. A study of past results and of the characteristics of the accounts receivable indicated that a suitably designed statistical sample would provide an estimate with the necessary precision from a substantially smaller number of accounts than the 15,000 accounts which had been selected previously. (Ref. 3, p. 298). Here, then, are two instances where use of statistical sampling techniques provided estimates with required precision, or led to conclusions with specified risks of incorrect conclusions, with significantly smaller samples than had previously been chosen.

Additional comments - It must not be thought, though, that statistical sample sizes will always be smaller than those conventionally taken in auditing. In some instances, the sample size required to provide specified precision or specified maximum risks of incorrect decisions may be greater than that previously used. Whether the statistically determined sample size is larger or smaller than the sample size conventionally used, benefits will accrue to the auditor from the use of statistical sampling procedures because he will then be able to evaluate and control the sampling errors involved in his various sample tests.

As stated previously, the application of statistical sampling techniques to auditing requires that the auditor formulate quantitative measures which meaningfully indicate the audit results in which he is interested. These formulations must take into account the over-all purposes of the audit as well as the specific purposes of each particular audit step. Furthermore, these formulations must consider the interrelationships which exist between various audit steps. Thus, it is not an easy task to formulate relevant quantitative measures of the audit results. Unless this is done, however, statistical sampling techniques

cannot be fruitfully employed to help the auditor in his sampling problems, such as the determinaticn of necessary sample sizes and the interpretation of sample results.

Applications in the Collection of Accounting Data

Accounting data are generally collected in order to aid management in the control of business operations, to analyze the past, and to plan for the future. Since they are not collected for their own sake, one must balance the cost of obtaining the data against their value. Often, management needs do not require perfectly "accurate" data; an estimate within a given percent of the "correct" amount is all that may be needed. In that case, the use of statistical sampling techniques may provide the information more cheaply than the 100 percent enumeration techniques which are generally employed. Furthermore, sample results may be more quickly available than data based upon complete enumerations. This would often be an important advantage since accounting data must be timely if they are to be valuable for control of current operations and for planning purposes.

A number of applications have been reported which illustrate these advantages of the use of statistical sampling methods for the collection of accounting data. A few of these will now be cited. The Chesapeake and Ohio Railroad has conducted an experiment to estimate inter-line charges on the basis of sampling. (Ref. 4). Railroad A and the Pere Marquette district of the C. and O. wished to ascertain the amount due the C. and O. on less-than-carload freight for a six-month period. The necessary information is available from the waybills, of which there were 23,000 for the six-month period. The computations required to ascertain the amount due the C. and O. on each waybill, however, are burdensome and expensive, and it was therefore decided to experiment with statistical sampling techniques in order to estimate the amount due the C. and O.

A preliminary investigation indicated that the efficiency of sampling would be greatly improved if the 23,000 waybills were first divided into separate groups, according to the amount of each waybill,and if each group were then sampled separately. For each group, a sampling ratio was determined; this ratio was larger for the waybills of large amounts than for the waybills of smaller amounts. Altogether, a sample of about 2,000 waybills - 9 percent of the total - was selected. From this sample, the total amount due the C. and O. was estimated by combining the sample results from the various waybill groups in an appropriate manner. Since this was an experiment, a 100 percent examination was also conducted in order that the sample result could be evaluated. The findings were as follows (Ref. 4, p. 63):

Total amount due C. and O. as determined from 100 percent study	$64,651
Total amount due C. and O. as estimated from sample	64,568
Difference	$ 83

The close correspondence between the sample estimate and the amount which was determined from a complete study of the 23,000 waybills should be noted. Two points should be stressed in this connection. The first one deals with the comparative costs of determining the amount due the C. and O. It was estimated that the sample estimate cost about $1,000

while the complete study cost about $5,000. Thus, it must be asked whether the additional accuracy achieved by making a complete study is worth its cost. Secondly, it should be pointed out that while the sampling error in this case favored Railroad A, the cumulative error over the long run will become relatively smaller and smaller if unbiased estimates are used in the sampling procedures.

An experiment on the inter-line settlements of commercial passenger revenue during a five-month period was also conducted. Railroads A and B and the Chesapeake district of the C. and O. were involved in this investigation; the results were as follows:

Railroad A	100% Examination	5% Sample	Difference Dollars	Difference Percent
(1) Total number of tickets	14,109			
(2) Total revenue	$325,600			
(3) C.& O. Portion of (2)	$212,164	$212,063	$101	0.05%
Railroad B				
(4) Total number of tickets	7,652			
(5) Total Revenue	$128,503			
(6) C.& O. Portion of (5)	$ 79,710	$ 80,057	$347	0.45%

Again the results indicate that the statistical sampling techniques provided estimates of high accuracy on the basis of only a small fraction of the items which would normally be included in accounting enumerations.

In another application, statistical sampling techniques have been used in order to estimate the base period cost of inventory with the LIFO valuation method. (Ref. 5). The company studied was a large manufacturer and supplier of machinery and equipment. A section of its inventory, valued at about $17 million in current costs, was included in the study program. This section contained about 250,000 items in over 100 different locations. A sample of about 25 percent of the items was selected, taking into account the different locations and product classes. Each inventory item selected was priced in terms of both current and base year costs by means of the regular inventory pricing procedures. From these data, the total value of the inventory at base year cost was estimated. A statistical evaluation of the precision of the sample estimate indicated that, with a 95 percent degree of assurance, it could be concluded that the base year cost valuation of the inventory at that time was somewhere between $11,138,000 ± .253 percent. In other words, the error range for the sample estimate with a high degree of assurance did not exceed ±.253 percent.

Since it was felt that an error range of ± 1 percent, with a reliability of 95 percent, was satisfactory enough for the purposes at hand, subsequent samples need include only about 4 percent of the inventory items. Here again, then, is an instance where accounting data of sufficient precision can be obtained by means of relatively small samples.

The Bell Telephone System has pioneered in the application of statistical sampling techniques to accounting records. In one instance, it

is necessary to ascertain periodically the distribution of telephones by type of apparatus; there are six such types. (Ref. 6, pp. 15-16). While a complete enumeration could be made from records maintained by the plant department, which show the type of apparatus at each customer location, the use of samples provides the information more quickly and cheaply, and almost as accurately. The telephones were first grouped into three classes - dial offices, non-dial offices, and private branch exchanges - and each class was then sampled separately. To obtain information of the distribution of telephones by type of apparatus through a complete examination would involve a costly job. The use of sampling in this instance was found to promise major savings, while providing estimates of sufficient precision.

In another instance, statistical sampling techniques have been used by several companies in the Bell Telephone System to determine the current average physical condition of their telephone plant. (Ref. 6, pp. 19-21). Not only was the sampling problem of great importance here, but non-sampling problems were also significant. Since the physical condition of the property had to be judged by inspectors, it was essential for obtaining reliable results that the inspectors be trained sufficiently so that their judgment be uniform. It was found that this uniformity of judgment could only be achieved by thorough training of the smallest practicable number of inspectors. This necessitated a relatively small sample since inspectors cannot be required to examine so many units of property that they are unable to examine adequately each unit inspected. Indeed, it was concluded in this instance that an accurate determination of the current average physical condition of plant could only be carried out by small samples since larger samples would involve human errors far outweighing the sampling errors that were encountered in this case. (Ref. 6, p. 20). While the precision of a sample required for submission to a public service commission is probably greater than that necessary for most other purposes, it was still found practicable to obtain an estimate of the average percent condition of the property as a whole with an error range of less than ± 1.0 percent, at a 99.5 percent level of assurance.

Many other illustrations of the application of statistical sampling techniques to accounting records in order to obtain timely and sufficiently precise data in an economical manner could be cited. Enough cases have been mentioned, however, to demonstrate the real usefulness of statistical sampling techniques in providing needed information from accounting records quickly and economically.

Applications in the Control of Clerical and Other Processes

Internal verification is carried on in many companies in order to assure management that the accounting operations are being carried out with reasonable accuracy. While it might be thought that the most effective way of carrying out this objective is to check each transaction handled by the accounting department, there are two major reasons why this might not be the case. In the first place, complete checking does not guarantee that all errors made will be found and corrected. Too often, it is simply assumed that inspectors are able to find all or most errors, even when such faith is actually not warranted. Furthermore, inspection alone can only discover errors after they have been committed. It would usually be more efficient if the errors could be prevented in the first place. A second reason why 100 percent inspection may not be the most effective method of assuring management of the accuracy of the

accounting operations is that the cost of 100 percent verification may exceed the benefits which it achieves. Accuracy, after all, is not desired for its own sake. It may pay a company to permit a small margin of errors in its accounting records rather than to try to eliminate them completely, as long as management can be assured that the extent of errors in the accounting records is reasonably small.

An interesting study of the verification of invoices received from vendors has been reported, where special attention was given to relating the cost of the verification procedure to the benefits obtained. (Ref.7). The study was made at the factory of an automobile manufacturer; it covered a period of seven months during which about 35,000 invoices were processed. Most of these were for small amounts; in fact, 80 percent of the invoices were for less than $500 and accounted for only 8.5 percent of the total dollar amount of all invoices. The same verification procedures for checking extensions, transportation charges, and quantities invoiced were applied to all invoices. The findings of this study cannot be presented in detail here; one phase of these findings will be sufficient to bring out some of their significance. Invoices for $500 and more, which constituted about 20 percent of all invoices, accounted for almost 80 percent of the net dollar amount of extension corrections and for almost 90 percent of the net dollar amount of transportation adjustments. Similarly, these large invoices were responsible for a major portion of the dollar amount of errors due to quantity adjustments.

Thus, this study indicated that in this particular case the great bulk of the dollar adjustments grew out of only a small proportion of the invoices. Under these circumstances, it may well be asked whether it is worth while to apply the same 100 percent verification procedures to all of the invoices since most of the verification expense arises from the examination of invoices which contribute only a small proportion of all dollar adjustments. Gregory, who conducted this study, concluded that the cost of processing each of the small invoices in this particular case exceeded the dollar amount of the errors discovered.

This type of analysis has indicated to a growing number of concerns that it may not be efficient to verify clerical operations on a 100 percent basis. Furthermore, the growth of the quality control philosophy has helped to emphasize in this area also that inspection results should be used for purposes of improving the quality of performance and not merely to find errors which have already been made. As a result, the use of acceptance sampling plans and quality control charts for controlling the accuracy of clerical processes and similar activities has become more and more widespread. For instance, at the 1954 Annual Convention of the American Society for Quality Control, Brinegar reported the successful application of acceptance sampling methods in order to control the accuracy of the inventory-taking at a large department store. (Ref. 8). Previously, a 100 percent check had been made of the work performed on the inventory; this nearly doubled the cost of the inventory and the time required. Since only a few of the teams taking the inventory were responsible for most of the errors made in any department and since large errors in price or quantity were infrequent, it was decided to use acceptance sampling methods for checking the accuracy of the inventory-taking. If the sample indicated that the work of a team is unsatisfactory, the previous work was checked 100 percent. It was found that the statistical sampling procedures really succeeded in separating the inefficient teams from the efficient ones. The savings in direct labor costs were substantial, and indirect savings resulted in addition

from the reduction in time required for taking the inventory. (Ref. 8 p. 315).

At the same meeting, Dalleck reported on the use of statistical control charts to control the accuracy of pricing airline tickets and the accuracy of the audit of these fare figures. (Ref. 9). It was found that statistical control charts and daily samples led to the prompt detection of problem situations - such as insufficient training, misinterpretation of new tariff regulations, volume peaks and mis-assignment of personnel - and consequently also to prompt remedial action.

Acceptance sampling plans are being used by the Standard Register Company for the verification of sales invoices.(Ref. 10). This company had been verifying all sales invoices before they were sent out. Despite this, some erroneous invoices were still being sent out. While the number of such invoices was not too large, the company was concerned about the cost of the verification procedure. A statistical sampling-verification procedure was, therefore, installed. Samples are taken at regular intervals. On the basis of the sample result, the group of invoices which is sampled is either accepted as being of satisfactory quality or inspected 100 percent if it is concluded that the group is of unsatisfactory quality. The statistical sampling program maintained the previous quality level, which was considered satisfactory, and did this at a saving of 47 percent in inspection time as compared with the earlier 100 percent verification procedure.

Groups of invoices which represent the work of several clerks are sampled in the Standard Register Company. In order that the sources of error can be located more precisely, however, records are kept for each clerk of the frequency and types of errors made by him as disclosed by the sampling inspection. These records are then used as an aid in determining appropriate remedial action. Such situations as improper or non-uniform training, faulty maintenance of source records, improper placement of personnel and inadequate methods or procedures have been brought to light by the statistical control procedures.

The Bell System for some time has been applying statistical sampling techniques in order to control clerical accuracy. (Ref. 6, pp. 9-12). One area where the application of statistical control techniques has been very satisfactory is the pricing of long-distance calls. This type of operation is repetitive; a large volume of work exists; errors can be clearly defined; and the work is completed at frequent intervals so that sampling of completed work permits prompt remedial action, when necessary.

A system of verification intervals has been used in conjunction with acceptance sampling plans in the Bell System in order to adapt the frequency of sampling to the quality record of the individual person. In controlling the accuracy of punching tabulating machine cards, for instance, the following system of verification intervals was used (Ref. 11, p. 8):

0
2 hours
1 day
1 week
1 month

Thus, the initial work should be verified at once by means of a sample. If it is acceptable, the work should be verified two hours later by means

of another sample. If it is still acceptable, the work should be sampled after a day has elapsed; and so on. If the sample at any time indicates that the work is not satisfactory, remedial action should be taken and the verification cycle starts all over again. In effect, then, this system provides relatively infrequent inspection for persons who are doing satisfactory work and frequent verification, as well as remedial action, for persons who are doing unsatisfactory work.

The use of a system of verification intervals with acceptance sampling plans not only provides control over clerical accuracy, but also locates a substantial portion of the errors which have been made. For example, as part of a series of tests by the Bell System, a system of verification intervals with statistical acceptance sampling plans was applied to control the accuracy of pricing long-distance calls. In addition to providing control over clerical accuracy, the statistical sampling procedure also located 56 percent of all errors made during this time by verifying only 12 percent of the work. (Ref. 6, p. 11).

The system of verification intervals need not be stated in terms of time but may, for instance, be expressed in terms of work units or assignments completed. The particular intervals to be employed depend upon such factors as the type of work examined, the importance of discovering unsatisfactory work, and the time and money available for inspection.

Statistical control techniques have also been used in order to control processes of concern to accountants other than clerical accounting operations. Noble, for instance, has reported the application of statistical control charts to the cost control of waste in a department converting rolls of paper into sheets. (Ref. 12). He has also cited the use of control charts in the analysis of daily variances in the number of container units produced. Statistical control techniques can also be applied to the analysis of costs for labor, materials, etc. Bicking, for instance, has described the application of a statistical control chart for analyzing the total production costs per 100 pounds of material produced over a period of time. (Ref. 13). This area of application of statistical control methods is still a relatively new one, but much work can be expected in this field in the near future. As accountants learn to appreciate the importance of keeping cost controls as close to each type of operations as possible and to do this on a frequent periodic basis in order that prompt remedial action can be taken when this is necessary and in order that the causes of difficulty can be located more easily, the use of statistical control techniques for cost control should spread rapidly.

<u>Summary</u>

The cases which have been cited in this paper should demonstrate that statistical sampling techniques can be of great help in auditing and in general accounting applications. In the area of auditing, the most immediate problem which must be faced before statistical sampling techniques can be fruitfully applied is the formulation of meaningful quantitative measures of the audit results in which the auditor is interested. Once this is done, statistical sampling techniques can help the auditor in determining necessary sample sizes and in interpreting the sample results.

In the collection of accounting data, statistical methods can often be of great value by providing information of required precision quickly and economically. Control of accuracy of clerical accounting work and of various types of costs may often be aided by statistical acceptance sampling plans, control charts, or other statistical methods which point out quickly when remedial action is required and which help to locate the sources of difficulty. Cooperation between accountants and statisticians will greatly aid the development of accounting uses of statistical sampling methods.

References

1. Vance, Lawrence L., and John Neter, Introduction to Statistical Sampling for Auditors and Accountants, to be published by John Wiley and Sons, Inc., New York.

2. Vance, Lawrence L., "An Experience with Small Random Samples in Auditing," The Accounting Review, October 1952, pp. 472-474.

3. Trueblood, Robert M., and R.M.Cyert, "Statistical Sampling Applied to Aging of Accounts Receivable," The Journal of Accountancy, March 1954, pp. 293-298.

4. "Can Scientific Sampling Techniques Be Used in Railroad Accounting?" Railway Age, June 9, 1952, pp. 61-64.

5. Rowles, Barry M., "Applications of Statistical Sampling Techniques to LIFO Inventory Valuation," The Accounting Review, April 1954, pp. 244-250.

6. Neter, John, "Some Applications of Statistics for Auditing," Journal of the American Statistical Association, March 1952, pp. 6-24.

7. Gregory, Robert H., "The Frequency and Importance of Errors in Invoices Received," Accounting Research, October 1952, pp. 332-339.

8. Brinegar, Claude S., "Department Store Uses of Statistical Quality Control," Quality Control Convention Papers 1954, American Society for Quality Control, pp. 311-315.

9. Dalleck, Winston C., "Quality Control at Work in Airline Accounting," Quality Control Convention Papers 1954, American Society for Quality Control, pp. 489-498.

10. Shartle, Richard B., "How a System of Continuous Scientific Sampling is Used to Control Accuracy in Large-Scale Invoicing," The Journal of Accountancy, August 1952, pp. 167-171.

11. Jones, Howard L., "Sampling Plans for Verifying Clerical Work," Industrial Quality Control, January 1947, pp. 5-11.

12. Noble, Carl E., "Statistical Cost Control in the Paper Industry," Quality Control Conference Papers 1952, American Society for Quality Control, pp. 9-15.

13. Bicking, Charles A., "The Application of Quality Control to Administrative Problems,"Industrial Quality Control,May 1950, pp. 21-25.

USE OF TASTE PANELS IN PRODUCT DEVELOPMENT

Gweneth J. Hedlund
General Mills, Inc., Research Laboratories

From the beginning of time man has been in search of foods which please the taste buds. People everywhere enjoy good food, i.e., food which has good flavor and good eating quality.

Therefore, if a food manufacturer is to be successful, it must establish taste and eating quality factors which have the most universal appeal. But how do we know when we have achieved this appeal? Unlike machines or non-edible products, flavor and eating quality can't be measured by objective means, such as micrometers, go no-go gages, chemical analyses, etc. Consequently, people in the food industries have had to look for reliable subjective means of measurement.

Fortunately, several methods have proved acceptable and can be found in the literature. While these sensory methods can be applied both in product development and control of product quality, we use them primarily in product development at the Research Laboratories. Therefore, this paper will deal principally with this phase.

At the outset, I might say that the objectives of our Sensory Panels are two fold -- 1) to give immediate product evaluation and guidance in product development and 2) to determine the stability or shelf life of products.

Before we go into some of the subjective means we use, let me give you a little background of our Taste Panel set-up, i.e., how we select our panels and the conditions under which we work.

Basically, we have three panels each of which is trained to test specific types of products. These panels are used primarily for determining whether or not differences exist. Probably one of the biggest problems in this work is to impress the people who are developing the product that the results should be considered in terms of differences, not consumer preferences. People tend to inject their personal likes and dislikes into product evaluation and must be constantly reminded to think in terms of differences only. To date Taste Panels can only tell us if differences exist, not what the consuming public likes. However, perhaps someday if Dr. Fox's (1) interesting findings can be adopted, we will be able to select our panels in such a way as to predict the public's reaction to the product flavor. As yet we are not that fortunate! In case you are not familiar with Dr. Fox's study, his taste "reactor tests" have indicated that 76% of the consumers seem to fall into 4 of 10 taste classifications.

Each of our panels is made up of fifteen to twenty people so that at least twelve to fourteen are available for each test. These people are engineers, chemists, food technicians, and some office personnel -- most of the panel members being associated with the type of product being tested. Much of their training has come from previous experience with the product. However, round table discussions are used to establish common terminology and new questionnaires to carry out our objectives. A running record of each tester is kept, that is, his deviation from the average of the group, in order to determine how critical the individual is, to eliminate the erratic testers, and to tell us if he is changing

CHART I

Sketch shows lay-out of the taste testing room. Entrance is through door at left. Sliding door separates food preparation and work area from the room where actual testing is done.

Versatility of the testing room is appreciated when conferences like this one are necessary. Booths fold against the wall, making space available for two collapsible aluminum tables. This arrangement will seat up to twelve people. Some of the testing work is most valuable when a group of people inspect samples and share opinions, as they are doing here.

his pattern or letting personal likes or dislikes enter into the picture.

Now let me show you our panel room. Chart I

This has been designed as a dual-purpose room. There are eight booths for individual testing which can be collapsed against the walls to make a larger conference room. Or tables can be put down the center aisle, then after each tester has recorded his independent opinions in the booth, he can turn around and discuss the test with the others in a round table discussion. Whenever necessary a variac lighting system is used to mask differences in color or appearance of samples. The outer room shown in the sketch serves as a preparation room and office space.

So much for the facilities used for testing and the make-up of our panels. Now we will discuss how we use them.

As I said before, basically, we use sensory panels for two types of evaluation, one for immediate product appraisal and the other for establishing the product's stability.

For immediate decisions, we use different methods of approach depending on the problem at hand. Some of these are: 1) matching samples to a known standard or control, 2) triangle tests, 3) paired comparisons, and 4) single product tests.

In matching an established control or standard, we use a questionnaire similar to that shown in Chart II. The testers have one sample marked control which they describe for flavor, texture or whatever characteristics are being tested. They are given two or three other coded samples and asked to test each in relation to the control indicating whether any difference exists, and if so, the degree of difference and to describe the difference.

There are advantages and disadvantages to this method. Some of the advantages are:

1. It is a good screening test, i.e., if one wishes to match a control and there are a large number of samples to be check, those that are grossly different can be quickly eliminated.

2. In a flavor test the descriptions indicate whether the difference is in flavor level or in character of flavor.

3. It indicates how large the difference is.

Some of its disadvantages are:

1. If the items are highly flavored, with so many samples there is a carry-over and/or a build up of flavor which tends to decrease the sensitivity of the individual. Also the order of tasting seems to influence the individual's reactions. By order of tasting, I mean whether the highest flavor level is tasted first, second, etc.

2. With a marked control, testers seem to look harder for differences and tend to pick up small differences which don't always exist. This has been verified when we have checked the panel by submitting identical samples. However, by analyzing the descrip-

CHART II

PRODUCT JUDGING SHEET

Describe control (flavor and texture) ______________________________

____

Do any of the test samples DIFFER from the control in flavor or texture? If so, indicate by checking opposite proper description in each column.

SAMPLE _____			SAMPLE _____	
Flavor	Texture		Flavor	Texture
_____	_____	No detectable difference or not certain	_____	_____
_____	_____	Definite, small difference	_____	_____
_____	_____	Definite, moderate difference	_____	_____
_____	_____	Definite, pronounced difference	_____	_____

Sample	FLAVOR DIFFERENCE DESCRIPTION	TEXTURE DIFFERENCE DESCRIPTION
_____	______________________	______________________
_____	______________________	______________________

Order of Tasting ________ ________ ________

Date: ________________ Judge: ________________

tions together with the degree of difference noted, we are able to determine whether or not a real difference exists. That is one of the reasons why it is important to have a large enough panel to be able to pick up these discrepancies.

Another variation of this form of test is to have one of the coded samples be the same as the control. The testers are to select the sample which is different from the control. This is a modification of the triangle type of test.

In triangle tests we use questionnaires similar to that shown in Chart III. Each tester is given three coded samples, two of which are identical and one different. He is asked to pick the odd sample for each characteristic being tested, to indicate how much it differs from the other two samples and to describe the difference. Incidentally, the fewer features testers have to evaluate in one test, the more sensitive the test will be. We usually test for only one and occasionally two features at a time.

One of the advantages of the triangle method is that it lends itself to statistical analysis. Fewer testers are required to pick up a statistically significant difference, if it exists, than to correctly match one of a pair of samples with a known control.

A limitation of the test is that only two different samples can be tested at a time. If there are a large number of samples, it requires a great deal of testing to evaluate all of them. Also, we have run into confused results due to carry-over of flavor when highly flavored items are being evaluated.

Following is an example of how we used these methods in one of our flavor studies. We had previously established a particular flavor balance as our standard for the product. Several suppliers had submitted some 15 to 20 samples as matches of our established standard. Our problem was to determine whether or not they were good matches and if not, why not. For the first three or four sessions we conducted round tables with about 10 of our most critical testers. The round table sessions checked 2 different samples against the control each time, weeding out those that were grossly different and checking whether the difference was in the level or in the character of the flavor.

Through these sessions we established the extent of our problem. In order to evaluate the differences statistically, we conducted triangle tests as just described, using one of our regular panels. After about 15 triangle tests we were able to list the flavorings which closely matched the standard in strength and character, as well as to describe how the other samples differed from the standard.

A paired comparison test is another method we use. Our questionnaires are similar to that shown in Chart IV. The tester is given two coded samples and is instructed to indicate whether or not a difference exists in any of the features being tested, the degree of difference and to describe the difference.

One of the biggest advantages of this method for us has been in the reduction of flavor interference or carry-over from one sample to the other. There also appears to be greater accuracy for the less experienced tester in that his flavor memory has to carry over only two samples

CHART III

PRODUCT JUDGING SHEET

Two of the samples are identical and one is different. You are to pick out the different one.

As nearly as possible, take the same amount of sample in each taste. Rinse your mouth and pause after each taste long enough to avoid interference between samples.

1. Indicate which sample is different and the degree of difference you noted. Use the following code for "degree of difference."

   0 -- None
   1 -- Possible slight difference, not certain
   2 -- Definite, small difference
   3 -- Definite, moderate difference
   4 -- Definite, marked difference

	Different Sample	Degree of Difference
________	________	________
________	________	________
________	________	________

2. How does this sample differ from the other samples?

________ ____________________

________ ____________________

________ ____________________

3. Which do you prefer and why?

________ ____________________

____

Order of testing samples: ____________________

Date: ____________ Judge: ____________

CHART IV

FLAVOR -- TEXTURE JUDGING SHEET

The purpose of this test is to determine whether or not there is any difference between the samples in flavor and/or texture.

1. Indicate whether or not these samples differ by checking opposite the proper description in each column.

FLAVOR	TEXTURE	
_____	_____	No detectable difference or not certain
_____	_____	Definite, small difference
_____	_____	Definite, moderate difference
_____	_____	Definite, pronounced difference

2. Compare flavor and texture of samples by describing differences.

Sample	Flavor	Texture
_____	____________________	____________________
	____________________	____________________
_____	____________________	____________________
	____________________	____________________

3. Which sample do you prefer and why?

____

Order of tasting: ____ ____

Date: ____________________ Judge: ____________________

CHART V

PRODUCT EVALUATION SHEET

After you have eaten the product, encircle the number that best describes your over-all reaction to it. Also, describe what you like or dislike about its flavor and eating quality (including texture).

CIRCLE THE NUMBER THAT BEST DESCRIBES YOUR REACTION

CODE ____

LIKE	
9--Extremely	9
8--Very much	8
7--Moderately	7
6--Slightly	6
NEUTRAL	
5--Neither like nor dislike	5
DISLIKE	
4--Slightly	4
3--Moderately	3
2--Very much	2
1--Extremely	1

COMMENTS

LIKES ----- Flavor ______________________________

____

Eating Quality ______________________________

____

DISLIKES -- Flavor ______________________________

____

Eating Quality ______________________________

____

Date: ____________________ Judge: ____________________

at a time. With an experienced panel, we have found that testing samples in pairs permits evaluation of several samples at one session, with a minimum of flavor fatigue.

Its disadvantage is that one cannot determine statistically whether the difference noted is significant. However, as I indicated before, we have found that analyzing the degree of difference noted in relation to descriptions of the difference has given us a pretty reliable picture.

When appraising a single product, we generally use a questionnaire similar to that in Chart V. We use this type primarily when we are just trying to get some indication of whether or not the product has merit, to get some degree of liking, and also to determine its weaknesses. Initially, a highly specialized panel of 8 or 10 people will evaluate it independently on this basis and then discuss it in a round table session giving suggestions to the individual working on the product and also giving him an opportunity to ask questions of the group.

We also use this type of questionnaire with our consumer type panel (people who are not experienced testers) to get some indication of product acceptance. Sometimes this group will evaluate two products comparatively on this scale, in which case the questionnaire will be the same except for allowance for rating and description of two samples. It should be emphasized, however, that this type of consumer panel testing is not intended to predict the consuming public's reaction to the product, but only to indicate whether the product is ready for consumer testing.

The development of one of our cake mixes may be cited as an example of the application of all these methods. Our objective was to determine what effect different blends of flavoring had on the over-all flavor. At the beginning, we ran into the problem of color differences. One sample tended to be darker than the other. When we ran pairs the tendency was to call the darker sample stronger in flavor. If a triangle test or a pair against a control was used, the odd sample could be selected by appearance. However, we overcame this problem by use of special lighting so that flavor could then be evaluated without bias. Chart VI shows the type of results obtained from one of these tests.

A pair against a control test was used, i.e., the marked control had Flavor X while one of the two coded samples had Flavor Y in it and the other was identical to the control. The testers were asked to select the sample which was different from the control. You will note that 11 of the 14 testers selected the correct sample which is significant. The difference was slight but definite and the descriptions indicated the sample with Flavor Y was milder. It was through this and the other methods of testing that we were able to assist in establishing the kind and level of flavoring to be used. After the formula seemed about right from the Laboratory's point of view, it was taken out to consumers for their reaction to it and to establish the tolerance of the product to different handling and equipment in homes.

It should again be emphasized that none of our panel testing is used to replace consumer testing, but is used only for guidance in product development and to tell us when we should go to the consumer for her evaluation.

CHART VI

FLAVOR X VS. FLAVOR Y -- FLAVOR DIFFERENCE TEST

Correct Sample Selection

Total Judges	14
Correct selection	11
Incorrect selection	1
No difference noted	2

Degree Of Difference Noted By Correct Judges

Total Correct Judges	11
Slight -- not certain	3
Definite -- slight	7
Definite -- moderate	-
Definite -- pronounced	1

Description Of Flavor Y Sample

Total Correct Judges	11
Stronger	1
Milder	10

These are just two examples showing our use of Taste Panels for immediate guidance in product work. Of course there are scores of others, each of which has had its own problems, however, the general pattern of testing has been quite similar.

Determining the products' stability or shelf-life is the other objective of our sensory panels. What effect do different types of packages, different ingredients, different processes, etc. have on the stability of a product? Which product is more stable and how long will it remain in good condition? Again flavor and eating quality are effected and this effect can only be measured in the laboratory subjectively by Taste Panels. Many of the same methods previously described or modifications thereof are used for these answers.

Before a storage test is set up to evaluate stability, samples being considered are checked by a special panel at a round table session to determine whether the product is ready for a storage test. The panel members individually describe the characteristics of each sample being studied. Then they discuss their findings and decide which ones to store, if any. The samples to be stored are then submitted to the regular panel to determine how the samples differ initially for reference in later comparisons. This will be done by paired comparisons.

Generally, our stability studies are based on three storage conditions -- for one the material is stored under accelerated conditions to measure the effect of high temperature in some states during the summer; for another the material is stored at Weather Room, where the temperature and humidity range widely over each 24 hours, to measure the effect of storage in warm humid areas; and the third condition is Room Temperature storage which of course varies with the season of year. However, it has a tempering influence on the more drastic and accelerated results which will be obtained from the other two conditions.

When we set up such a test, some of each variable being tested is placed under each of the three storage conditions. The rest of the packaged samples are put in freezer storage. Samples are then removed from the cold room at set intervals and placed under the different storage conditions. We schedule our first checks on the stored material as follows:

For high temperature we check material which is 0, 4 and 6 weeks old.

For Weather Room the check is 0, 4 and 8 weeks.

For Room Temperature the check is 0, 6 and 12 weeks.

This may vary with the past experience of the product being tested, however, we set the first checks at intervals which we are sure will detect when changes begin to show. The material called "0 weeks old" has been taken from cold storage and serves as the control each time.

Chart VII shows the type of questionnaire used. The control is marked and the other two samples are coded. The panel knows it is a storage test but does not know the age or storage condition being tested. They describe the flavor and texture of the control sample and indicate the degree of difference of each sample from the control and describe this difference. They also record their opinion of the condition of the sample due to storage.

CHART VII

JUDGING SHEET

Describe control (flavor and texture) ______________________________

____

Do any of the test samples DIFFER from the control in flavor or texture? If so, indicate by checking opposite proper description in each column.

SAMPLE ____			SAMPLE ____	
Flavor	Texture		Flavor	Texture
____	____	No detectable difference or not certain	____	____
____	____	Definite, small difference	____	____
____	____	Definite, moderate difference	____	____
____	____	Definite, pronounced difference	____	____

Sample	FLAVOR DIFFERENCE Description	TEXTURE DIFFERENCE Description	If any off aroma noted, describe.
____	____________	____________	____________
____	____________	____________	____________

CONDITION OF SAMPLES DUE TO STORAGE

Sample	No detectable difference or not certain	Edible, but definite change		Not edible
		Small change	Moderate change	
____	____________	____________	____________	____________
____	____________	____________	____________	____________

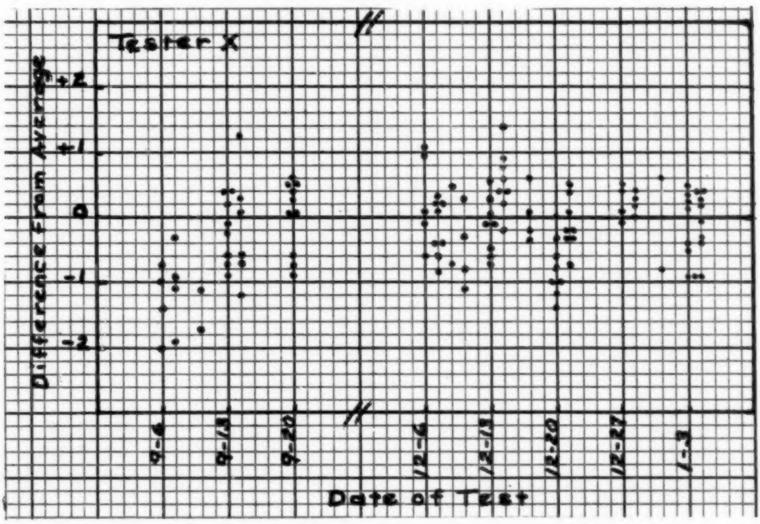
Order of tasting: ____ ____ ____

Date: ________________ Judge: ________________

The description of the control alerts us if any changes have taken place in it.

The extent to which each sample has changed helps to check accuracy of the panel members because any changes noted should be in relation to the age of the samples. A record is kept in control chart form which shows each individual's deviation from the average of the group at each check. These are in chronological order and give us a good appraisal of the individual's performance. See Chart VIII.

CHART VIII



The time for the next check is scheduled in accordance with the results of the current test. For instance, if in the Weather Room check of samples with 0, 4 and 8 weeks storage, no changes are noted at 4 weeks but differences are noted in the 8 week samples, the next check may be scheduled for 0, 6 and 10 weeks. This checking of products at scheduled intervals continues until the samples are out of condition or considered inedible. Incidentally, our panel is far more critical in this respect than are consumers.

To complete the story of which product is more stable, paired comparisons are made between sets of samples of the same age.

This method is very flexible and reduces the amount of testing necessary. Now it can be done with 6 to 8 tests, while previously in checking at routine intervals it required 2 to 3 times as many tests. Besides the advantage of fewer tests, checking two age periods at the same time against a control serves as a stabilizing factor and gives a better overall trend picture of the changes that occur with time. Although checking against a control tends to give a more critical picture of the condition

of the sample, the evaluation of the paired samples tends to temper the ratings and give a more realistic picture as it would appear to the consumer.

The following charts show how the results of this type of test are recorded to give a running story as the test progresses.

CHART IX

Storage Test #851

	Product A			
			Degree of	
Storage Condition	Diff.*	Cond.**	Staleness	Rancidity
HIGH TEMPERATURE				
4 Weeks	.9	1.8	Indication	---
6 Weeks	1.3	1.7	Slight	---
WEATHER ROOM				
6 Weeks	1.1	1.8	Slight	---
10 Weeks	1.5	1.7	Slight	---
11 Weeks	1.6	1.8	Slight	---
15 Weeks	1.6	1.6	Slight to moderate	---
ROOM TEMPERATURE				
9 Weeks	.9	1.8	Slight	---
13 Weeks	1.7	1.7	Slight	Indication
15 Weeks	1.2	1.7	Slight	---
17 Weeks	2.1	1.4	Slight	Indication
19 Weeks	2.1	1.2	Moderate	Slight
23 Weeks	2.9	.9	Moderate	Pronounced

*Difference From Control
- 0--No change
- 1--Possible slight difference--not certain
- 2--Definite small difference
- 3--Definite moderate difference
- 4--Definite pronounced difference

**Condition Due To Storage
- 2--Relatively little change
- 1--Edible, but definite change
- 0--Not edible

Chart IX shows how the sample changes from its control under each storage condition. The first column indicates the extent to which the sample has changed from the control and the next column the condition of the sample due to storage. The last two columns were set up to catch the description of flavor changes. These vary with each test, but in this case we were particularly interested in finding the stability as regards staleness and rancidity.

CHART X

Storage Test #851

Storage Condition	Diff.*	Product A vs. Product B — Condition & Description Of Difference** A	B
HIGH TEMPERATURE			
6 Weeks	2.8	1.8--Good	.6--Mod. stale, ranc.
7 Weeks	3.0	1.6--Sl. stale	.4--Rancid
WEATHER ROOM			
10 Weeks	1.9	1.9--Good, sl. flat	1.5--Stale, ind. odd & ranc.
11 Weeks	2.3	1.7--Good, ind. flat	1.0--More stale; ranc.
15 Weeks	2.6	1.6	.4--More stale & ranc.
ROOM TEMPERATURE			
13 Weeks	2.4	1.8--Good	1.3--Mod. stale, ranc.
15 Weeks	2.9	1.6--Flat	.3--Stale, more ranc.

*Difference Between Samples
- 0--No difference
- 1--Possible slight difference--not certain
- 2--Definite small difference
- 3--Definite moderate difference
- 4--Definite pronounced difference

**Condition Due To Storage
- 2--Relatively little change
- 1--Edible, but definite change
- 0--Not edible

Chart X shows the results of paired comparisons. The first column shows the degree of difference between the two samples. The next two columns show the condition of each sample due to storage and descriptions of how they differ. In this particular test it can be noted that A appears to be considerably more stable than B.

CHART XI

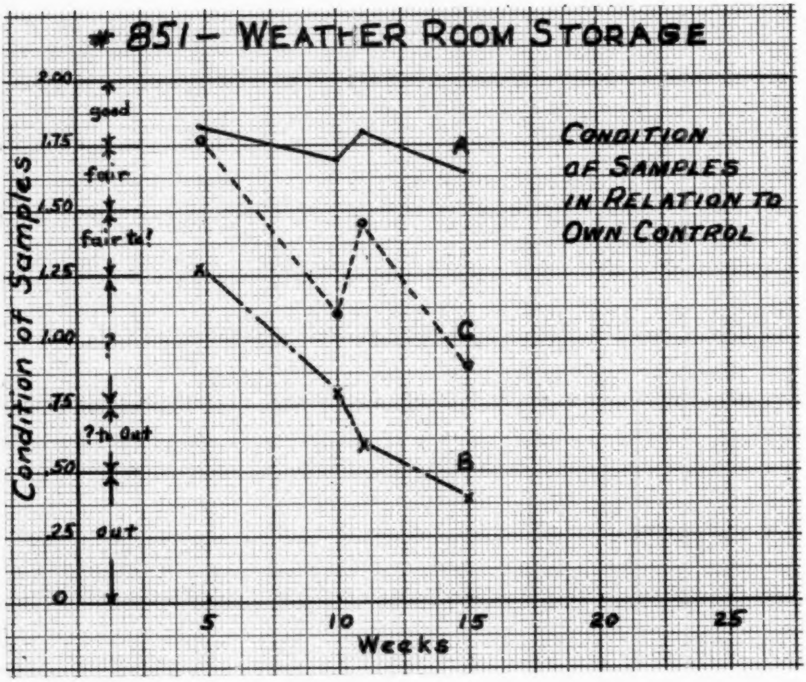


Chart XI is typical of the type of chart we use to give a quick graphical summary of a study. This particular graph shows the condition of each sample due to storage in relation to its own fresh control. At the same time it also gives a comparative story between the samples.

To recapitulate, we at General Mills Research use Taste Panels or Sensory testing with two purposes in mind:

1. To get product evaluation for immediate guidance in development or improvement; and

2. To determine the stability of products.

We use many of the techniques described in literature, or modifications of them, such as, matching control, triangle, paired, and single product tests. The kind used depends on the problem at hand, since no one type is best.

Although time and volume of work limits the extent to which we can train special panels and conduct duplicate tests, our methods have given

sufficiently accurate information to show trends and give reliable guidance to those working on the products. Probably one of the most important parts of the analysis of any test for us is the careful examination of comments. They are often the deciding factor as to whether or not a difference actually exists. Whenever the results are not conclusive, the test is repeated or another that might serve better is substituted.

The help that government and college laboratories[2,3,4] have given us in devising and verifying sensory testing methods has enabled us to keep up with the new developments. We use these new methods or variations of them whenever possible. We haven't arrived at our present setup overnight nor do we expect to continue testing in exactly the same manner as in the past. Each project or group of tests points up good and bad features and teaches us something new which we try to incorporate in the next project. This is the reason why Taste Panel work is so challenging and interesting.

REFERENCES

1. Fox, Arthur F. "New Direction To Taste Testing", Food Engineering, January, 1955.

2. Boggs, Mildred M. and Hanson, Helen L. "Analysis of Foods By Sensory Difference Tests", Advances in Food Research, Vol. 2.

3. "Sensory Methods For Measuring Differences in Food Quality", Agriculture Information Bulletin No. 34.

4. "Food Acceptance Testing Methodology, A Symposium" edited by David R. Peryam, Francis J. Pilgrim and Martin S. Peterson, Quartermaster Food and Container Institute.

MANAGEMENT OF THE QUALITY CONTROL FUNCTION

Dr. A. V. Feigenbaum and William J. Masser
General Electric Company

Within the last decade quality activities have led us to believe that quality control must have a double-barreled objective.

First - To maintain and progressively improve product quality.

Second - To realize a substantial reduction in the costs for improving and maintaining quality.

There is a best way to meet these objectives, through a complete and positive quality program which starts with the design of the product and ends only when the product has been placed in the hands of a customer who remains satisfied with it.

Quality activities begin the moment the salesman and the customer begin discussing specifications and the customer's quality needs.

It continues on when the design engineer translates these needs into definite dimensions, tolerances and finish requirements.

Quality activities progress to the manufacturing engineer who assigns equipment and methods of performing the needed operations as well as the materials required.

The quality program is influenced by purchasing in choosing, contracting with and retaining vendors for parts and materials. Manufacturing supervision and shop operations have a major quality responsibility during parts making, sub-assembly, and final assembly.

Also very important is the quality effect during mechanical inspection and functional test in checking conformance to specification. Shipping has its influence in the caliber of the packaging and transportation.

Successful "make it right the first time" quality control thus involves a broad quality cycle where quality is recognized as everybody's responsibility. Being recognized as everybody's job, quality may very well become <u>nobody's</u> job!

With the all importance of quality and quality costs, it requires that these responsibilities be <u>buttressed</u> and <u>serviced</u> by the operation of a recognized well-organized function whose <u>only</u> specialization is product quality at minimum quality costs. (Chart 1)

This quality control component has two basic responsibilities.

First - To provide <u>quality assurance</u> for the departments' products.

Second - To assist in assuring <u>minimum quality costs</u> for these products.

The quality assurance responsibility is achieved by carrying on the <u>necessary inspection and testing</u> both to establish that the products shipped are of the quality specified by engineering and also to <u>feed back</u>

facts for preventing the production of unsatisfactory quality in the future.

Responsibility number two—assistance to obtain minimum quality costs—is achieved through Quality Control Engineering and Equipment Design work. This consists of planning and analysis effort to help assure that quality will be right before production starts. The feedback cycle (Chart 2) becomes the lifeline of quality control.

The quality control activities that are carried on to meet these two responsibilities are four in number:

1. New Design Control
2. Incoming-material Control
3. Product Control
4. Special Process Studies

(Chart 3)

New Design Control is a preproduction quality control activity where quality level and capability information is supplied to the design engineer for his use in establishing the best and most practical product specifications. It involves conducting and analyzing inspections, tests and studies on samples of pilot runs so that the feedback of this information will insure quality trouble-free tools and fixtures. It requires the determination and planning of the inspection and test operations that are required to assure optimum product quality during production. It involves also the design of inspection and testing equipment and the means of interpreting the inspection and test data that results from use of this equipment.

Incoming-Material Control is a manufacturing quality control activity carried on while vendors' materials and parts are purchased, received, examined and released for use.

.The activity involves analysis of the quality levels and capabilities of potential vendor sources of supply in cooperation with purchasing. It requires checking out samples and first parts received and establishing vendor certification and rating routines. At the same time, it must assure the receipt of quality products.

Product Control is the measurement of parts or assemblies at the point of production to discover quickly errors in manufacture so as to be able to initiate immediate corrective action. Operator education, preventive maintenance programs, process sampling techniques and process capability analyses are a few of the tools employed in product control.

Special Process Studies are the intense critical surveys and tests condusted to locate the specific cause of quality problems and to provide information that will lead to improving product quality. Of particular importance is the fact that information learned on today's quality problems can be applied to tomorrow's production to prevent poor performance and costly delays.

In each of the above four activities, statistical quality control methods play a useful role. Frequency distribution data, machine and process control charts, and sampling plans are developed for use by inspector, production operator, foremen, vendors and others to help them

do their quality jobs more easily and of course more economically.

The quality control program described gains definite results: systematic improvements in product quality levels, shortened manufacturing cycles, reduced spoilage and rework costs, reduced quality costs. It also helps to promote a keen sense of quality mindedness throughout the shops and offices.

QUALITY CONTROL ORGANIZATION

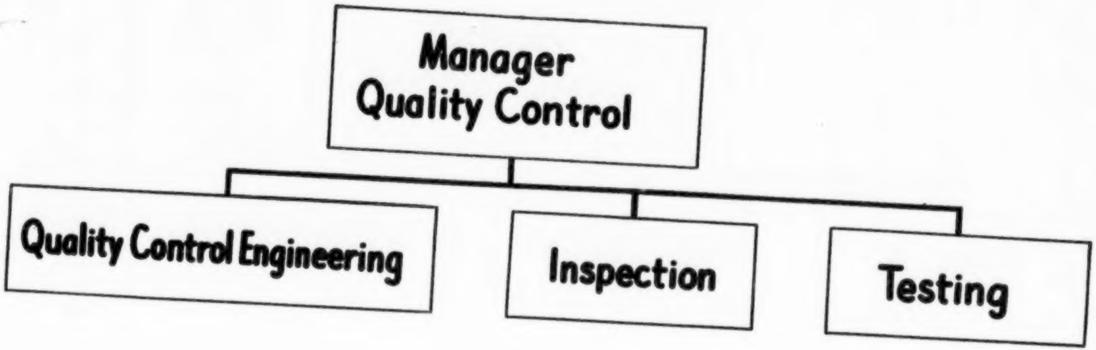


CHART 1

QUALITY CONTROL FEED-BACK CYCLE

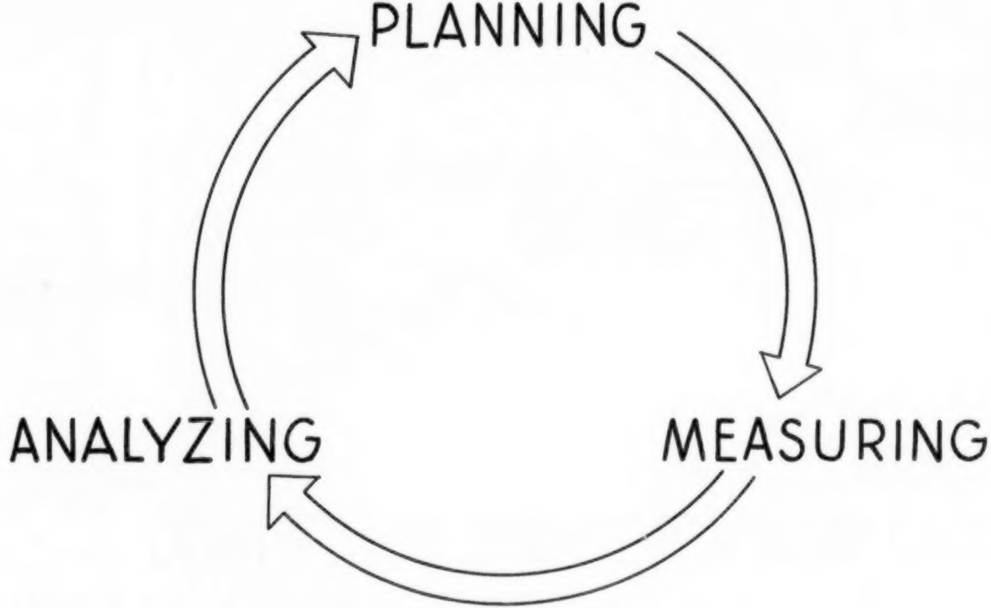


CHART 2

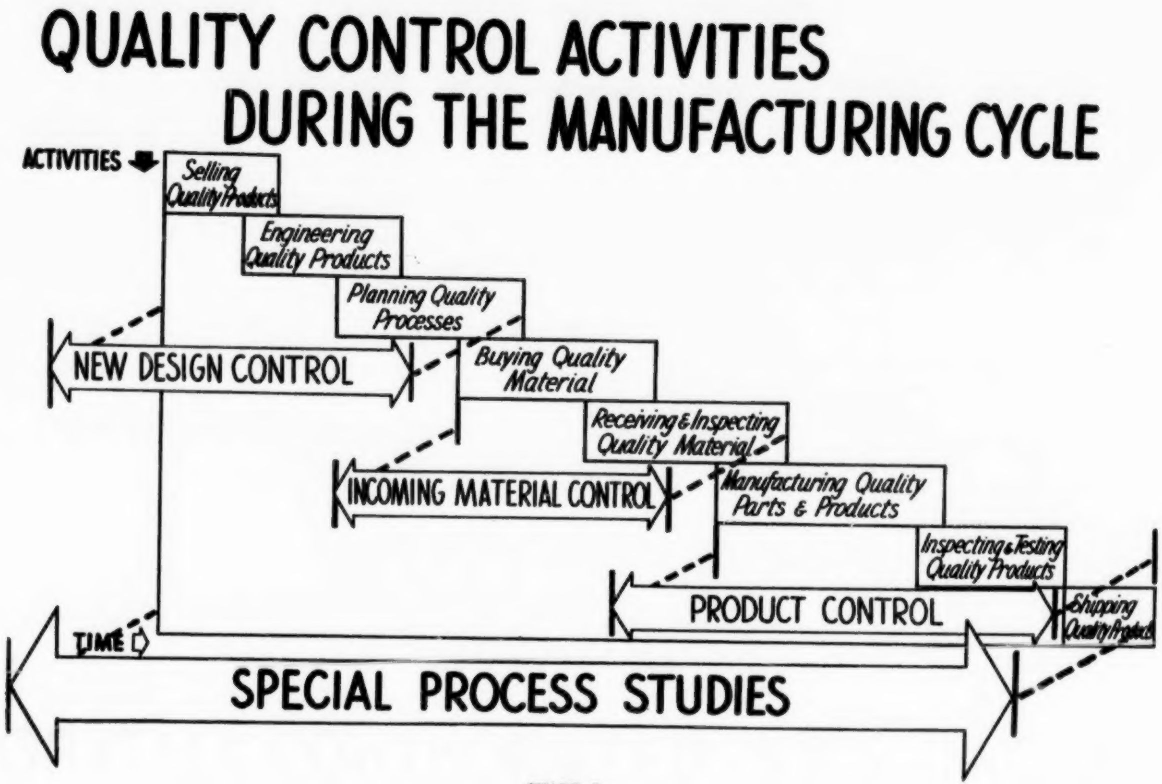

QUALITY CONTROL ACTIVITIES
DURING THE MANUFACTURING CYCLE
ACTIVITIES
Selling Quality Products
Engineering Quality Products
Planning Quality Processes
Buying Quality Material
Receiving & Inspecting Quality Material
Manufacturing Quality Parts & Products
Inspecting & Testing Quality Products
Shipping Quality Products
NEW DESIGN CONTROL
INCOMING MATERIAL CONTROL
PRODUCT CONTROL
TIME
SPECIAL PROCESS STUDIES

CHART 3

THE RELATION BETWEEN STANDARDS AND QUALITY CONTROL

G. F. Hussey, Jr.

The Third National Conference on Standards had as its theme - STANDARDS — ENGINEERING TOOLS FOR INDUSTRY. Today we purpose to examine the relation between two important engineering tools for industry - standards and quality control.

Perhaps the earliest relation occurs in the well-known American War Standards Z1.1, 1.2 and 1.3-1942 which, for the first time, made readily available to the quality control practitioner the basic instructions for the application of the control chart method. Developed by a highly competent committee, drawn very largely from your own number and brought out under the press of war conditions at the urgent request of the War Department, these American War Standards have stood the test of time so well that there has so far been no move for even a revision.

In setting up standards for the dimensional, physical or chemical characteristics of a particular product there is recognition that variations are inevitable and that accordingly tolerances must be applied.

Some years ago when I was Chief of the Armor and Projectile Section of the Bureau of Ordnance, I had an experience in trying to purchase in accordance with a Federal Specification for balloon cloth in which there were no tolerances on weight or tensile strength. There was much to-do when the samples failed on the tensile test. Buying against the contractor's account produced no better material. The result - which may have been foreseen - was the necessity for rejecting all bids and revising the specification to provide tolerances.

The methods used in setting up a quality control chart of themselves serve in their initial application of any job to determine the natural limits which represent the variations to be expected in the quality of the product of any given process. If those natural limits are not adequate for the standards which must be set, then it becomes apparent that a change must be made in the process, in the equipment, or - as a result of a re-examination - in the standard itself. Once the natural limits are determined and found to be satisfactory the process may be started. So long as the natural limits are not transgressed the process is said to be in control and it is then working at its maximum capability.

With a recognition of this relationship between the standards and the ability to control quality according to these standards, there should be an end to the warfare that so often occurs between engineering and production. Too often the tolerances indicated on the drawings were set with little more for a background than a pious hope that they could be met, and the inability of the shop with existing equipment to meet the tolerances resulted in recriminations between shop and engineering. With a soundly developed standard the application of the quality control chart assures the output of the best possible product with the equipment available and random variations in product showing up outside limits on the chart indicate faults

in material, tool, or basic equipment. So long as the process remains in control there is no point in wasting time seeking to improve it because within the natural limits the machine or the process is doing its best.

Perhaps some of you are familiar with the initials JIC. They stand for the Joint Industry Conference. This is a more or less amorphous group composed of representatives from the automotive industry and from their suppliers in the machine tool and industrial equipment groups with their suppliers in turn of electrical, hydraulic and pneumatic equipment. The JIC standards for electrical, pneumatic and hydraulic equipment on tools and other industrial appliances for the automotive and high production industries lay particular stress on the avoidance of down time. In a mass production industry where each tool is working close to its maximum capacity, the failure of a single piece of equipment can throw out of balance the whole production scheme. The quality then which is built into this equipment through compliance with the JIC standards provides a means of quality control of an element seldom measured; that is, the continuity of production time.

An important and by no means side element in the JIC standards lies in their safety provisions. These are directed not merely at the operation of the machine itself, but at the means for machine maintenance so that when and if repairs or servicing are necessary, they can be accomplished in a minimum of time with the least possible hazard to the serviceman concerned. That being the case, the JIC group have recently produced a pamphlet on safety interpretation of their electrical standards. Profusely illustrated it is an excellent handbook for the manufacturer, for the installation man and for the maintenance man in the plants.

For many years now there have been reports of industrial accidents made in accordance with the American Standard for Reporting Industrial Accidents. The fruits of these reports have had a vital bearing on the premiums which companies have had to pay for their casualty insurance protection. Within the last few weeks this standard has been revised by the committee in charge and approved by the American Standards Association under the title of American Standard Method of Recording Work Injury Experience. There is an important distinction between the two titles because the new one recognizes the concentration of the standards on the effects of accidents on individuals. There is thus in the recording of such injuries a type of quality control which is directed at the prevention of injuries to individuals.

Like the character in a classic French play who had been speaking prose all his life without knowing it, the safety engineers have long been practitioners in quality control in their charts recording accidents and their effects on workers. They have in general had to work with incomplete data because many accidents never come to their attention. It would seem to be of importance, equal to that of preventing injuries, that there should be attention directed to the prevention of accidents. For example, if a truck transporting materials within a plant loses a part of its load, it becomes a matter of interest to the current standard on work injuries only if somebody is hurt, whereas the prevention of the accident which fortuitously hurt no one should be of equal interest to the safety engineer of the plant.

The problem facing the safety engineer is too much complete data on all mishaps in order that his quality control chart may point out to him the areas most in need of his attention as well as those where the most immediate effective results can be obtained. The numerous American Safety Standards are the tools against which the safety engineer applies quality control methods in a field at once humanitarian, conserving, and of great economic importance. Standards and quality control thus form an effective team in an area far removed from those usually associated with statistical quality control.

In the development of standards, whether within a technical society or a trade association or by an ASA Sectional Committee, an important consideration is - can the quality be controlled within the limits contemplated? A corollary to this is - has the quality been specified higher than necessary for the ultimate performance of the finished product? It is, of course, true that standards have often been used as a means for raising quality, but this is not necessarily the primary objective of standardization.

The human tendency to set limits closer than necessary has probably cost industry - at home and abroad - vast sums of money. To be realistic in this area requires a bold approach and a determination to waste neither time nor money in seeking a precision which is not requisite to the end product. The development of a standard on this principle calls for a competent understanding of the effects of tightening or relaxing tolerances on the producability of components and their final assembly into the finished product, as well as the accumulation of these effects on the performance of the end product. Time and effort spent in this stage of standards development can pay big dividends in the production of acceptable parts and assemblies with controlled quality.

Basic to all of this discussion is a recognition of just what is quality and how it enters into the picture. Without attempting a dictionary definition (of which there are many with none appearing to bear specifically on our present problem), it seems to me that quality in any item is the degree of approach to the relative perfection called for by the standard which is pertinent to the case. Thus, if the end use will be served by holding a dimension within plus or minus one-half an inch and the dimension is so held, it would seem that the required quality has been obtained and has been controlled in accordance with the needs. On the other hand, should a tolerance of plus or minus five thousandths be essential to the successful operation of the final assembly and this tolerance is met, then again the quality is all that is desired and is controlled in accordance with the standard. Standards and quality control then go hand in hand for quality control must have a standard as a starting point and the degree of success in meeting the standard is evidenced by the chart results in the application of quality control.

CRITERIA FOR SELECTION OF ATTRIBUTES
SAMPLING ACCEPTANCE PLANS

Herbert Arkin
Bernard M. Baruch
School of Business and Public Administration
City College of New York

The use of attributes sampling inspection plans is widespread in industry, arising in large part from the example set by government practices such as those exemplified by Military Standard 105A.

However, while commonly used sampling plans are generally well founded upon adequate statistical theory, the selection of the particular plan used in a given situation is oft determined by "practical" considerations rather than through an adequate statistical approach.

While it is conceded that the selection of a particular plan in a given situation is a function of the management objective which initiated the inspection system, nevertheless, the selection of a plan to meet that objective should not be fixed by so called "practical" considerations, but rather the ability of the plan to accomplish that which management has in mind.

Some of the "practical" considerations that often decide the plan selected include:

1. Cost of inspection.
2. Psychological disadvantage of a small sample.
3. Availability and characteristics of sampling plans in published tables.
4. Legal considerations.
5. Attitude of the vendor.

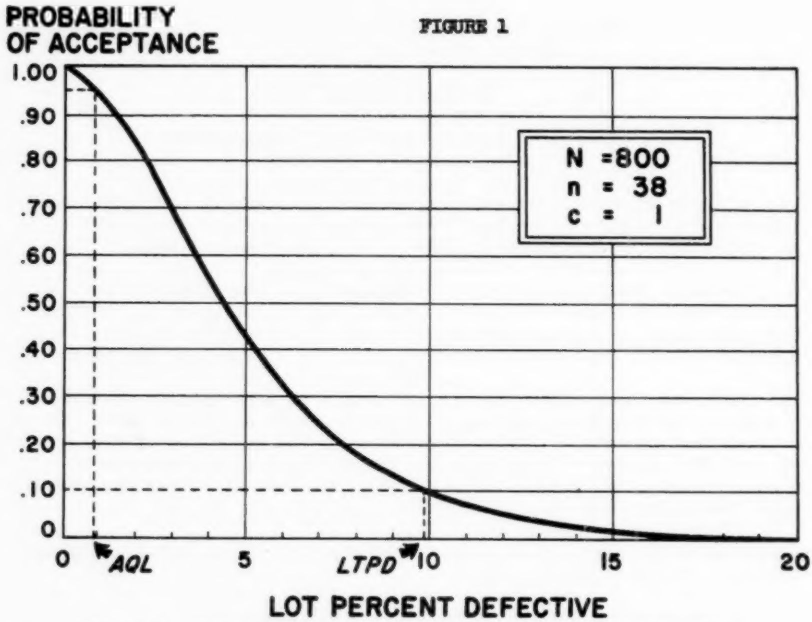
However, the so called "practical" considerations should neither dictate nor be the prime factor in selection of a sampling plan. Such a method of selection defeats the very objective of the plan. The inspecting company merely fools itself. A plan that costs less but does not accomplish management objectives is a waste of time and money and might better not be used at all.

The selection of a sampling inspection plan should be based upon the characteristics of the sampling plan considered and the relation of these characteristics to the management objectives to be attained.

The characteristics of a sampling inspection plan are indicated by its operating characteristic curve. This operating characteristic curve which is unique to each sampling plan indicates the probability of accepting a lot submitted to the sampling plan when that lot has various levels of quality (percent defective). A graphic representation of a single sampling plan[1] is shown in figure 1 on the following page.

[1] For purposes of illustration, attention will be continued to consideration of single sampling plans but these observations apply in equal force to double and sequential plans.

FIGURE 1



The O. C. curve of the sampling plan is in turn fixed by the basic facts of the plan, namely, lot size (N), sample size (n) and the acceptance number (c). There is only one curve possible when these 3 factors are specified.

Examination of the O. C. curve (figure 1) for the plan N = 800, n = 38 and c = 1 indicates various probabilities of acceptance of submitted lots when they contain various percentage of defectives[1].

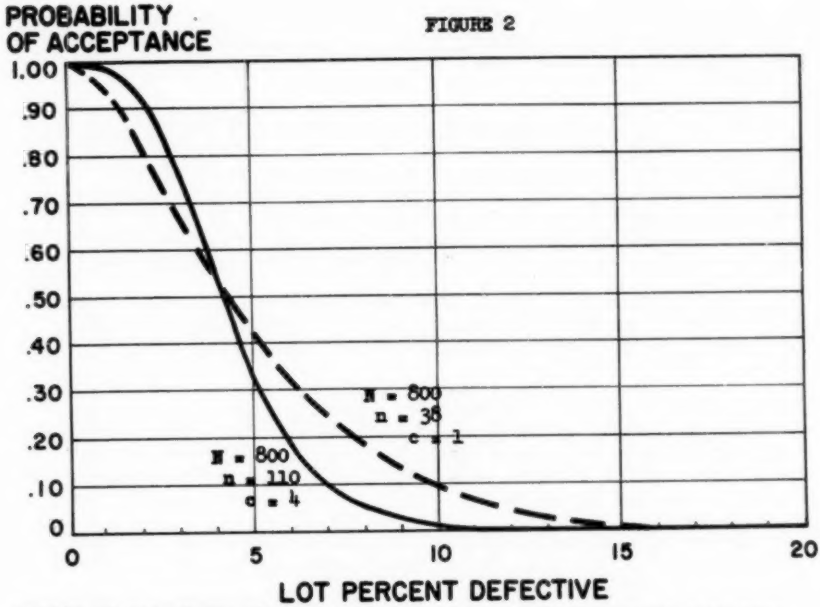
On the other hand other plans such as that in figure 2 (N = 800, n = 110, c = 4) will have different O. C. curves if any one of these 3 factors are varied.

The selection of the particular plan to be used becomes a problem of matching the abilities of a sampling plan to reject bad lots and accept good ones as specified by the O. C. curve and the stated or conceived management objectives to be accomplished by the inspection.

It is important to note that while several such curves may cross at one point on the graph, or in other words may have the same proba-

[1] Attention is invited to the fact that these probabilities, as all probabilities, are "long run" values and do not mean that every finite group of lots will develop exactly the percentages of acceptance indicated by the curve. For a finite group of lots the actual number of lots accepted while tending to approach the values of the O. C. curve will in turn be dictated by probability.

FIGURE 2



bility of acceptance when a given lot of a specified percent defective is submitted to the test, the balance of the curve or the probabilities of acceptance of other incoming fraction defectives will be unique.

However, the selection of a plan by qualified personnel who base their selections on the characteristics of the plan usually is based on a single point on the curve. For instance, the general approach has been to select a plan either according to the Lot Tolerance Percent Defective (LTPD or p_t), the Acceptable Quality Level (AQL) or the Average Outgoing Quality Level (AOQL). All of these are single points on a curve which is characteristic of the plan.

This situation has arisen largely as an outcome of the physical limitations in preparing sampling plan tables such as the Dodge-Romig or Military Standard 105A tables.

It has not been found feasible to prepare tables which show all or a large number of points on the O. C. curves for many plans and as a result only one (in the case of the Dodge-Romig one on the O. C. curve and one on the AOQ curve) point is indicated to reference to the plan. In addition, that one point is selected in accordance with the definition of the criteria peculiar to that table.

The Lot Tolerance Fraction Defective is one point on the O. C. curve indicating the incoming percent defective associated with some small value of probability of acceptance. The Dodge-Romig tables through wide usage has "frozen" this probability (consumer's risk) at 10%. Little consideration is given to the fact that the 10% probability is an arbitrarily selected value and will not serve all purposes.

As to AQL, although past usage has caused a 95% probability of acceptance to be used as a criterion, because of the SRG tables, the Military Standard 105A tables use a varying and unstated probability (85 to 99%).

However, the mere use of these values in fixing the LTPD and AQL points "straight jackets" the use of the plans in the tables. Such definitions may not be correct for the problem at hand. Perhaps another probability might be more appropriate for defining LTPD instead of the customary 10% or the 95% for the AQL. If so, it is necessary to revert to the O. C. curve. Of course, the graphs in Military Standard 105A are a valuable contribution in this direction.

Various other individual points have been used to define a plan. For instance, the value of percent defective associated with a 50% probability acceptance has been called the "indifference" point.

More recently[1] attempts have been made to develop a single criteria based on the slope of the O. C. curve for different plans by stating the tangent at the point of inflection. This rather technical approach is again based on a single O. C. curve point, that at the inflection. Other attempts were made by developing a ratio between the AQL value and the LTPD.

However, no single point tells the entire story. It is a consideration of the <u>whole</u> O. C. curve that dictates the utility of the plan for a given purpose.

For instance, let it be assumed that the management of a company has found that with respect to a particular component material, examination of economic considerations such as the cost of rejected end products, production stoppage, consumer ill will and complaints, indicates that lots of this material containing 10% or more defectives are undersirable. If it is further assumed that the lot size is 800, a plan defined as $n = 38$, $c = 1$ will give the LTPD with a 10% consumer's risk.

However, this concept of LTPD merely indicates that lots of 10.0% or more defective will be <u>rejected</u> at least <u>10%</u> of the time.

But lots as high as 12% defective will be accepted about 4.4% of the time as shown by the O. C. curve (see figure 1). A probability as high as 4.4% of acceptance of bad lots may be found undesirable. It is then necessary to redefine LTPD so that the probability of rejecting lots more than 10% defective is much higher. This cannot be obtained directly from available sampling tables.

On the other hand, any plan based on a sample also runs a risk of rejecting good lots, a situation which may be unfair to the producer. For instance, for the above plan while the 95% AQL is 1%, lots only 2% defective will be rejected 17% of the time--patently unfair to the producer.

[1] A Method of Discrimination for Single and Double Sampling O. C. Curves, Utilizing the Tangent at the Point of Inflection, Bush, N, Leonard E J and Marchant MQM, Engineering Agency, Special Report, Army Chemical Center, 2 October 1953.

It will be necessary to secure a sampling plan which will weed out undesirable lots to a required degree but raise the probability of acceptance of better lots (lower fraction defective) to a more reasonable level.

This examination of the entire O. C. curve discloses this as an inadequate sampling plan even though the Dodge-Romig LTPD is 7% and the 95% AQL is 2%. The other probabilities cause the trouble.

A different plan is essential. For instance, the plan N = 800, n = 110 and c = 4 will give better rejection of higher fraction defective lots (probability is 0.8% at 10% defective) and reject fewer lots at 2% defective (probability 6%).

However, this will increase the sample size. Further the Dodge-Romig LTPD no longer is the same. Decision as to the feasibility of this new approach then rests on the "practical" considerations mentioned above.

The problem of selecting a plan then resolves itself into a determination of the objective and examination of <u>all</u> points on the O. C. curve to determine the desirability of alternate possible sampling plans.

This makes the use of published sampling tables based on a single value of limited utility. The tables can be used as a guide in selecting possible alternate plans but the whole O. C. curve must be available in order to make a final determination.

It is seldom that management, sampling tables to the contrary, will be satisfied to accept even one out of 10 lots of a value stated as unsatisfactory. The requirements will probably be stated in terms of rejection of a much higher probability. Sampling plan tables do not provide this alternative.

It may be feasible to develop more adequate tables for this purpose. Perhaps a book of graphs of numerous O. C. curves will serve the purpose. If some interested person has the facilities available, it would be a valuable contribution to statistical quality control to publish such a document.

It is hoped that a better understanding of this problem will result in more adequate applications of sampling inspection method involving a more critical evaluation of the plans and less acceptance of plans on faith. It is only through elimination of such uncritical acceptance of plans which really do not meet desired objectives that sampling inspection can find the really widespread acceptance it deserves.

STATISTICAL DESIGNS FOR TASTE TEST PANELS

Ralph Allan Bradley
The Virginia Polytechnic Institute
and Rutgers University

PANEL SELECTION

The selection of a taste panel depends on its purpose and different methods of selection will be used for different types of panels. The expert panel is used for research work involved with the detection of differences or for the maintenance of quality control. Taste panels for consumer acceptance and for quality evaluation will generally be obtained in different ways.

Taste panels for consumer acceptance are usually large and untrained. The sampling problems are those of sample surveys and panel members should be chosen with a view to obtaining a sample of consumers well representative of the population of consumers. The selection of the panel is best accomplished following accepted methods of probability sampling. The same considerations regarding stratification and resultant methods of estimation as are required in public opinion polls will be necessary. On the selection of a panel for consumer acceptance studies, the large size of the panel or sample, its dispersion throughout a population, and the lack of control over the conditions under which food samples are considered dictate the use of very simple experimental designs. It will usually be necessary to use a simple questionnaire and to request little more than simple preference statements on comparisons of only two or three food items. On most consumer acceptance panels, panel members are requested only to select the preferred of two items. The statistical analyses are simple and based on binomial distributions. All too often only an over-all percentage preference for one of the two items is recorded but better techniques with stratified samples would use methods of estimation for stratified samples to estimate both percentage preferences and the variancesof the percentages. In the latter method, estimates and their variances depend on known population characteristics, usually population sizes in each stratum.

In considering quality evaluation, we may think of taste testing as only one phase of a more elaborate evaluation procedure. Composite quality scores consist of weighted averages of a variety of determinations. This kind of quality evaluation is used in certain United States Standards for Grades. The tasting is done by a very small number of official graders. These graders are selected for their knowledge of the pertinent production field and receive intensive training with a view to standardization of concepts of quality. Interest is in an absolute taste score and not in comparative scores for several test items as is usually the case in other types of panel testing.

Taste panels for the detection of differences and for quality control are usually selected by a screening process from a larger group of potential panel members. The two types of panels have very similar purposes but the panel for quality control is likely to restrict is attention to difference judgments on taste of only one or a very limited number of products while the panel for the detection of differences is used for research purposes for tests on a much wider range of food items. The selection of the panel for quality control should be based on tests involving the product to be controlled while the taste panel for

differences may be selected, for example, on the basis of ability to differentiate between small differences in basic tastes. There is however some disagreement as to whether skill at differentiating between basic tests is sufficiently well correlated with distinguishing ability in the complex taste characteristics of actual foods.

The selection of panel members for difference or quality control panels is usually based on repeated triangle or duo-trio tests. In the first, the individual is requested to select the odd sample from a set of three wherein two samples are identical; in the second, the individual selects the sample from a pair that matches a specified third sample. Usual procedures depend on giving repeated triangle tests (ten to fifteen) to potential panel members and on then selecting the required number of 'best' tasters. The number of triangle tests presented is usually small enough that there is a real possibility of selecting some poor tasters for the panel and of rejecting some very good ones.

Alternative procedures (3) to the use of fixed numbers of triangle tests for the selection of 'best' tasters use sequential methods to select tasters of satisfactory discriminating abilities. For specified abilities, the sequential method will on the average require fewer triangle tests per individual than the fixed sample size method for specified risks of accepting poor tasters or rejecting good ones. The general reaction to the use of sequential methods is that they require too many tests. This only indicates that methods customarily used do not sufficiently well control Type I and Type II errors. The sequential method has the advantage that it focuses attention on these risks of incorrect decisions and a sufficient number of acceptable panel members may be selected without screening the entire group of potential panel members. The essential difficulty entering is in defining acceptable ability. If this is set too high, it may not be possible to find enough acceptable people and it may require an impractical number of tests. But the average numbers of trials for a decision can be worked out in advance. Sequential methods are recommended in that they are generally more efficient and in that they automatically focus attention on the risks of accepting poor tasters and of rejecting good ones. Details for the use of the well known Wald method of sequential analysis and of a newer method of Rao are given in the reference along with examples of their uses.

SCORING SCALES

A number of appeals for the standardization of scoring methods in taste panel work have recently appeared. This seems to be related to the use of scales with a uniform number of points and uniformity in descriptive terms describing the scale points or scores. The idea in this desire for standardization seems to be that there will then be a real possibility of comparing the research of different laboratories and research centers. This seems to overlook the possibility, and we believe it to be a real one, that different taste panels (and even different members of one panel) differ in their interpretations of scoring systems.

The units of measurement in most experimentation are clearly defined and follow naturally from the formulation of the problem. In taste testing there is no natural unit of measurement in terms of which subjective decisions and judgments may be recorded. The determination of a scale should follow from the assignment of scores to two distinct standard food samples in the sense that two points on a line should be sufficient to determine its origin and scale. In practice the use of

standards does not work because it is possible to taste only small numbers of samples at one time due to fatigue factors and the standards must also be tasted.

When a scoring scale is defined, difficulties enter owing to non-uniformity of scoring, lack of consistency of individual judges, lack of agreement among judges, the effect of order of presentation of samples, psychophysical adaption, and doubts as to the appropriateness of standard methods of statistical analysis.

ANALYSIS OF VARIANCE

The general assumptions of analysis of variance are:

(i) Observations are independent in probability,

(ii) Observations come from normal populations,

(iii) Error variances are homogeneous, and

(iv) Treatment and environmental effects are additive.

Taste fatigue may introduce departures from (i) and (iii); a discrete scoring scale without the use of standards usually leads to departures from (iii) and, with the effect of adaption, from (iv); the discreteness of a scale always leads to violation of (ii) although this may not be too serious. Difficulty with the assumptions for the valid application of analysis of variance may not lead to incorrect decisions but it is at least difficult to defend the validity of conclusions based on analysis of variance.

Two methods of analysis of variance have been devised with some of the problems of taste testing in mind. Scheffé (14) has developed an analysis of variance for paired comparisons wherein food items are scored in incomplete blocks of size two. (The use of small incomplete blocks is indicated by fatigue factors in taste testing.) The essentially new features of Scheffé's procedure is that the effects of order of presentation of samples to judges may be measured and allowed for in treatment comparisons. Scheffé has given a comprehensive report on this research along with examples and we refer the reader to the reference.

Calvin (9), with a method based on scores like Scheffés, considered doubly balanced incomplete block designs and inserted additional parameters into the linear model of analysis of variance to allow for adaption, the effect of the presence of one treatment on another. He called these additional parameters 'correlations'.

A balanced incomplete block design is one in which pairs of treatments appear equally often in incomplete blocks. The usual linear model is that

$$[1] \quad y_{hi} = n_{hi}(\mu + \beta_h + \tau_i + \epsilon_{hi})$$

where

y_{hi} is an observation on treatment i in block h,

n_{hi} = 1 if treatment i occurs in block h
= 0 otherwise,

μ represents the average level of scoring,

β_h represents the effect of block h, perhaps due to the taster doing the scoring, the time of day, etc.,

τ_i is the effect of treatment i, and

ε_{hi} is a random error assumed to be independent of other random errors and to have a normal distribution with zero mean and constant variance.

Calvin modified [1] in the form

$$[2] \quad y_{hi} = n_{hi}(\mu + \beta_h + \tau_i + \sum_{j \neq i} n_{hj} m_{ij} \alpha_{ij} + \varepsilon_{hi})$$

where the symbols have the same meanings as in [1] and α_{ij} is the effect of the presence of treatment j on the observation on treatment i in block h. If treatment j does not appear in block h, $n_{hj} = 0$ by its definition. m_{ij} has the value 1 if $i < j$ and -1 if $i > j$. This only means that the effect of treatment i on treatment j is the negative of the effect of treatment j on treatment i. Calvin's model is an attempt to allow for the effect of adaption in taste testing experiments. He found that he could estimate the additional parameters in [2] easily (but with considerable additional arithmetic in examples) if only the balanced incomplete block design were doubly balanced. A doubly balanced incomplete block design is defined as a balanced design where in addition all triplets of treatments appear in incomplete blocks an equal number of times.

We leave the readers to consult Calvin's paper for the details of numerical analysis. We have used his method with ranking within incomplete blocks and by then transforming the ranks to scores using Table XX (Scores for Ordinal or Ranked Data) given by Fisher and Yates (11). It is our experience that, with ranking, the correlation effects measured by α_{ij} are not important and that it is sufficient to use simply balanced incomplete block designs. This is reasonable unless one believes that α_{ik} and α_{jk} are of such magnitude as to reverse the ranking of y_{ih} and y_{jh}. That is, unless the presence of treatment k in the h^{th} block with treatments i and j has so great an effect as to change the apparent order of treatments i and j. If ranking is used, we would not then include the correlation effects of [2] in our model but rather use the simpler form [1]. It is perhaps well to check our experience as Calvin's design is used with other food items.

PAIRED COMPARISONS

Our considerations up to this point suggest that desirable experimental designs for sensory difference tests are those which employ ranking and incomplete blocks with small numbers of treatments in a block. A design that has these characteristics is based on what is known as the method of paired comparisons. The present author has assisted in the development of a new method of analysis for paired comparisons which was devised for taste testing. The wide use of our methods in taste testing experiments in industry, agricultural research and in studies involving such diverse subjects as photography and radioactive trace elements seems to indicate the acceptability of the procedures.

Consider t treatments, $T_1,\ldots,T_t$, in an experiment involving paired comparisons. A repetition of the experiment is defined to be a set of the $\frac{1}{2}$ t(t-1) incomplete blocks of size two possible taking all pairs of

treatments. Parameters, $\pi_1, \ldots, \pi_t, \pi_i \geq 0, \sum_i \pi_i = 1$, were postulated and associated with the treatments. These parameters are supposed to be such that the probability that T_i ranks above T_j in an incomplete block with treatments T_i and T_j is $P(T_i > T_j) = \pi_i/(\pi_i + \pi_j)$. The null hypothesis of treatment equality is then expressed as $H_0: \pi_i = 1/t$, $i=1,\ldots,t$. Test procedures for various alternatives to H_0 and with other null hypotheses are set forth in (2) and (15). Variations in the specification of treatment parameters are considered and it is sometimes possible to divide the repetitions into groups (perhaps by judges, time, batches, etc.) and to test for group by treatment interactions.

In applying a statistical technique, it is desirable to check that the mathematical model is appropriate. Methods for doing this for paired comparisons have been outlined (5) and some experimental data shown. In addition, Hopkins (12) obtained larger samples with a view to checking the appropriateness of our models. The properties of the method of paired comparisons have been investigated (8) but this work has not yet been published. Abelson, with the present author, (1) considered the superimposition of factorial arrangements of treatment on paired comparisons but the useful results obtained are limited to the two by two factorial.

RANKING METHODS

We have recently given applications of the more useful ranking methods in a two-part paper (6,7) and we shall only point out here that the methods most appropriate to taste testing are those that utilize ranking within blocks. For two treatments, the appropriate method is the sign test, and, for k treatment problems, one should use the method of concordance with complete blocks of size k if k is not too large. For incomplete block designs, the generalization of the concordance method given by Durbin (10) is available. His method is not well known but is illustrated using taste test data in (7).

Durbin considered n treatments in balanced incomplete blocks of size k, each treatment ranked m times in the experiment, and with $\lambda = m(k-1)/(n-1)$, the number of times pairs of treatments appear in incomplete blocks. The analysis depends on computing the rank totals for each of the n treatments and S, the sum of squares of deviations of treatment rank totals about their average, $m(k+1)/2$. The concordance coefficient is $W = 12S/\lambda^2 n(n^2-1)$ and $\lambda(n^2-1)W/(k+1)$ has approximately, for moderately large values of m, a chi square distribution with (n-1) degrees of freedom. Durbin also gives an F- approximation that is somewhat superior to the use of chi square.

The chief criticism levelled at the use of ranks is related to the supposed loss of efficiency in comparison with the use of numerical observations. This criticism does not appear to be valid in taste testing in that scoring scales are difficult to devise and use and in addition tend more to classify responses than to represent a true measuring system. The added difficulty of expressing judgments in terms of scores in comparison with that of assigning ranks also suggests that sample sizes may be increased with ranking without increasing the total time of experimentation.

SUMMARY

We have tried to note the more recent developments in statistical methods of taste testing. It was not our intention to discuss the

details of their applications. We have given references to the pertinent papers and in addition an extensive classified bibliography of taste testing methods is given in (3). Savage (13) has recently provided a large bibliography on rank-order and nonparametric methods and this also will be useful.

1. R. M. Abelson and R. A. Bradley. A 2 X 2 Factorial with Paired Comparisons. Biometrics 10, 487-502, 1954.

2. R. A. Bradley and M. E. Terry. The Rank Analysis of Incomplete Block Designs, I. The Method of Paired Comparisons. Biometrika 39, 324-345, 1952.

3. R. A. Bradley. Some Statistical Methods in Taste Testing and Quality Evaluation. Biometrics 9, 22-38, 1953.

4. R. A. Bradley. The Rank Analysis of Incomplete Block Designs, II. Additional Tables for the Method of Paired Comparisons. Biometrika 41, 502-537, 1954.

5. R. A. Bradley. Incomplete Block Rank Analysis: On the Appropriateness of the Model for a Method of Paired Comparisons. Biometrics 10, 375-390, 1954.

6. R. A. Bradley. Some Notes on the Theory and Application of Rank Order Statistics, Part I. Industrial Quality Control, 11, 12-16 1955.

7. R. A. Bradley. Some Notes on the Theory and Application of Rank Order Statistics, Part II. Industrial Quality Control, 11, 1955.

8. R. A. Bradley. The Rank Analysis of Incomplete Block Designs, III. Some Large Sample Results on Estimation and Power for a Method of Paired Comparisons. Submitted to Biometrika.

9. L. D. Calvin. Doubly Balanced Incomplete Block Designs for Experiments in which the Treatment Effects are Correlated. Biometrics 10, 61-88, 1954.

10. J. Durbin. Incomplete Blocks in Ranking Experiments. Br. Journal of Psychology. Stat. Section, 4, 1951.

11. R. A. Fisher and F. Yates. Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd, Edinburgh, 1948.

12. J. W. Hopkins. Incomplete Block Rank Analysis: Some Taste Test Results. Biometrics 10, 391-399, 1954.

13. I. R. Savage. Bibliography of Nonparametric Statistics and Related Topics. Journal American Statistical Association 48, 844-906, 1953.

14. H. Scheffé. An Analysis of Variance for Paired Comparisons. Journal American Statistical Association 47, 381-400, 1952.

15. M. E. Terry, R. A. Bradley and L. L. Davis. New Designs and Techniques for Organoleptic Testing. Food Tech. 6, 250-254, 1952.

VARIABLES THEORY - THE CONTROL CHART FOR AVERAGE AND RANGE

Irvin W. Schoeninger
The Centralab Division of Globe-Union Inc.
Milwaukee, Wisconsin

Introduction

From the description of frequency distributions given in the previous paper, it is evident that they are very useful for determining machine and process capability based on accumulated data, particularly when the order of production is not known. For operations requiring relatively quick decisions about resetting or allowing a machine or process to run, frequency distribution analysis is often too slow and because of the larger number of measurements required may be too costly. Fortunately, there is another set of statistical tools known which provide the means for making decisions which need often to be made very quickly.

Description of the Control Chart for Variables

The Control Chart for Variables is one of the most potent tools for controlling quality during actual manufacture. It is based on the fact that variation will follow a stable pattern as long as the system of chance causes remains the same, and is designed to detect the presence of "assignable" causes of variation (unstable patterns of variation) both as to process level (centering) and variability (spread).

Once a stable system of chance causes is established, the limits for the "only to be expected" pattern of variation can be determined. These control limits are placed symmetrically above and below the grand average of the sample averages ($\bar{\bar{X}}$) and the sample ranges ($\bar{R}$) at a distance such that when a point exceeds the limits, the odds are approximately 300 to 1 that the occurrence was not due to chance, but to an assignable cause. This type of control chart is designated as an $\bar{X}$ and R chart and is particularly adaptable where economy of effort is important and where a continuous record of performance is desired. It is a valuable instrument for the diagnosis of quality problems and the routine detection of sources of trouble.

Definition of Terms

Before proceeding to elaborate on the "why" and the "how" of $\bar{X}$ and R charts, it will be helpful to define the following symbols and terms which are commonly used:

X	Measurement of one item in a sample.
n	The number of items in a sample.
$\bar{X}$	Average of a sample.
$\bar{\bar{X}}$	Grand average of the averages of a series of samples.
R	Spread or range in a sample. (max.-min. measurement).
$\bar{R}$	Average of the ranges of a series of samples.

A_2	A constant used for control limits for sample averages.
D_3, D_4	Constants used to obtain control limits for sample ranges.
σ_X	Standard deviation of a sample (statistical measure of variation of individual values). A measure of spread.
$\sigma_{\bar{X}}$	Standard deviation of sample averages.
d_2	A constant which relates the range to the standard deviation. This constant is used in obtaining the spread of the individual values of a process.
$UCL_{\bar{X}}$ $LCL_{\bar{X}}$	Upper and lower control limits for sample averages.
UCL_R LCL_R	Upper and lower control limits for sample ranges.

Universe-Total group of units in which we are interested.

$\bar{X}'$	Average of the universe.
σ_X'	Standard deviation of the universe.

Relationship between the Universe ($\bar{X}' \pm 3\sigma_X'$) and Sample Means ($\bar{X}$)

In order to understand how the control charts for $\bar{X}$ and R with their respective control limits serve as a guide in determining if the established pattern of variation is maintained throughout the production run, we must establish certain basic relationships. The first one is between the universe ($\bar{X}' \pm 3\sigma_X'$) and the sample averages, and the second one is between the universe variation (σ_X') and the average sample range ($\bar{R}$).

It is a well established fact that sample averages will themselves form a frequency distribution which exhibits the characteristic pattern of all frequency distributions, i.e., a central tendency and variation in either direction. The average $\bar{\bar{X}}$ of such a distribution tends to approximate $\bar{X}'$ the average of the universe. The spread of this distribution of $\bar{X}$ values depends on the universe spread (σ_X') and also on the sample size (n). This is so because statistical theory tells us that in the long run the standard deviation of the frequency distribution of $\bar{X}$ values may be expressed as $\sigma_{\bar{X}} = \sigma_X'/\sqrt{n}$. We are also told that if the universe is normal, the expected frequency distribution of the $\bar{X}$ values will also be normal. Therefore, since any normal distribution can be completely specified if its average and standard deviation are known, this means that in sampling from such a normal distribution of averages, the complete picture of the expected pattern of variation is given by the expression $\bar{\bar{X}} \pm 3\sigma_{\bar{X}}$. This is shown graphically in Figure 1 for averages of samples of 4, 9 and 16.

Furthermore, even though the distribution in the universe of individuals is not normal, the distribution of the $\bar{X}$ values tend to be close enough to normal to permit being specified by the values of $\bar{\bar{X}}$ and $\sigma_{\bar{X}}$ as previously stated. In support of this, Dr. Walter A. Shewhart(1) shows examples of distribution of $\bar{X}$ values from a normal, rectangular and triangular universe which indicate a close fit to the normal curve in all three instances.

Relationship between Universe Standard Deviation(σ_X') and Range ($\bar{R}$)

Thus far we have considered only the behavior of sample averages which we now recognize as a measure of the central tendency of the universe from which they are selected. Another important consideration is the measure of universe variation, σ_X', as related to the average sample range $\bar{R}$. Statistical theory has established the expected relation between these as follows: $\sigma_X' = \bar{R}/d_2$. The d_2 ratio factor for various sample sizes will be found in Table 1 with other constants for control charts. We have also to consider the fact that variation occurs in the value of sample range. It seems that at every turn we are faced with the necessity of determining the limits of variation of one sort or another. We have the measure of universe variation (σ_X'), then the measure of variation for sample averages ($\sigma_{\bar{X}}$), and now a measure of the variation of the universe standard deviation (σ_{σ_X}') which, even though no simple formula exists for it, may be expressed as σ_R. General Leslie E. Simon (2) in his presentation of sampling by variables, comments on this situation with the following quotation from De Morgan's "A Budget of Paradoxes":

Great fleas have little fleas upon their backs to bite 'em,
And little fleas have lesser fleas, and so ad infinitum.

We are not an authority on fleas and so cannot verify De Morgan's statement, but a somewhat parallel idea certainly applies to the statistical theory of distribution. Just as each universe has an average and standard deviation, so does each distribution of sample averages and ranges have an average and standard deviation.

Control Limits for $\bar{X}$ and R

It has been previously stated that the limit of the expected pattern of variation for sample averages is $\bar{\bar{X}} \pm 3\sigma_{\bar{X}}$. This then becomes the basic formula for the control limits for averages and can be more simply expressed as $\bar{\bar{X}} \pm A_2\bar{R}$. The constant A_2 is one of a series which are listed in Table 1 to be used for calculating control limits for average and range for various sample sizes. These were developed during World War II as the result of a study to simplify the computation leading to the various standard deviation calculations. The formula for control limits for range is $\bar{R} \pm 3\sigma_R$, which in simpler terms results in a lower control limit expressed as $D_3\bar{R}$ and an upper control limit, $D_4\bar{R}$.

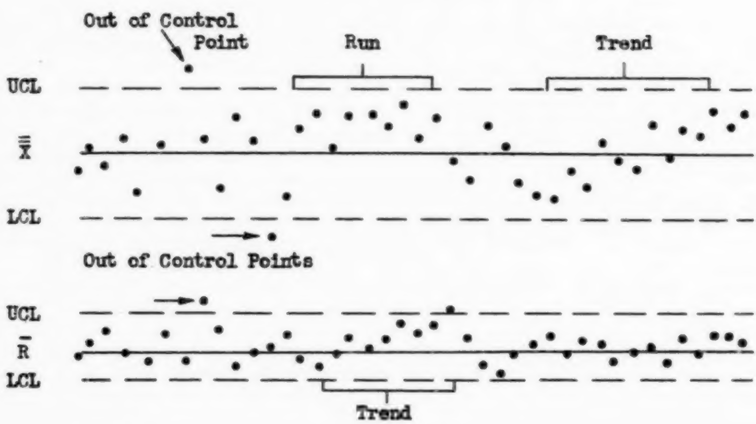
In the actual construction of a control chart, 15-25 sample averages and ranges are calculated and the points plotted on cross section paper, the averages in the upper half of the chart and the ranges in the lower half. The grand average $\bar{\bar{X}}$ of sample averages is determined as well as the grand average of sample ranges $\bar{R}$. A solid line for $\bar{\bar{X}}$ and $\bar{R}$ are drawn in their respective portions of the control chart, the control limits for $\bar{X}$ are placed at a distance $A_2\bar{R}$ above and below the average line and the control limits for range $D_3\bar{R}$ and $D_4\bar{R}$ are placed below and above the average range line. These then define, respectively, the boundaries of the chance fluctuation of sample averages and sample ranges and are called action limits, because if they are exceeded corrective action is indicated. The portion of the chart for averages relates to the centering of the process and for ranges relates to the variability or spread of the process.

Why 3-Sigma Control Limits?

Trouble in the sense of the existence of assignable causes of variation in quality is a common state of affairs in manufacturing. Under such circumstances, it does not pay to hunt for trouble unless there is a strong indication that it really exists. The real basis for control chart limits is experience with those limits which strike an economic balance between two kinds of error, looking for trouble which does not exist and not recognizing it when it is present. Such experience indicates the desirability of 3-sigma limits, for when closer limits such as 2-sigma are used, the control chart often gives indication of assignable causes of variation which cannot be found, whereas when 3-sigma limits are used and points fall out of control, the cause of the trouble will usually be found by careful investigation.

Interpretation of the Control Chart

Even though a control chart gives evidence of satisfactory control, it is important to look for sequences in control chart data. Considerable work has been done by mathematicians on the development of various types of statistical tests based on the theory of runs. In order to detect shifts in a process average in manufacturing, the most practical plan is to use a few simple rules that depend only on extreme runs. The following are suggested by E. L. Grant(3): Whenever 7 out of 7, 10 out of 11, 12 out of 14, 14 out of 17, or 16 out of 20 successive points on a control chart are on the same side of the central line, it may be assumed that a change has occurred in the process. These sequences will occur by chance more frequently than will a point fall outside of 3-sigma limits and for this reason are a somewhat less reliable basis for hunting trouble, although very useful to detect a process shift when points do not fall out of the control band. The tendency for trends to develop should also be observed, because this may be an indication of tool wear or machine drift. If this is so, then frequent resetting may be necessary if specifications are narrow, or modified control limits may be used if the specifications are wide enough to permit. The following sketch may be helpful in interpreting the control chart.



Modified Control Limits

The control chart thus far has been used as a means of determining whether a stable pattern of variation continues to be produced. The pattern of variation may be a wide one or a narrow one depending on the process. Since most processes are required to meet some specifications, it is desirable to be able to learn from control chart data how the actual process compares with the desired. It should be remembered that specifications are frequently man made to fit a particular end use whereas the control chart shows the variation caused by the irreducible factors in the process. The specification generally relates to individual values, whereas the control chart is based on sample averages. In order to obtain the total actual variation of individual values for comparison with specification, the following formula may be used but only when the process is in control: estimated limits for individuals equals $\bar{\bar{X}} \pm 3\bar{R}/d_2$. The limits thus calculated may be compared with the desired specifications to determine conformance.

When the estimated process limits for individuals exceed specification, then an investigation of the machine or process should be made to determine if corrective action can be taken to reduce the spread. If not, then the least amount exceeding specification will occur if the process is centered on the specification mid-point. When, however, the process spread is less than allowed by the specifications, modified control limits may be used. Through the use of these limits, full advantage may be taken of the broader specification, and the process need not be held on exact center. The process average may be permitted to shift somewhat with respect to specification mid-point. The basic customary formulas for modified control limits are:

$$\text{Upper Modified Control Limit} = \text{Upper spec limit} - (3\sigma_X' - 3\sigma_{\bar{X}})$$

$$\text{or, more simply, UMCL} = \text{Upper spec limit} - (3\bar{R}/d_2 - A_2\bar{R})$$

$$\text{Lower Modified Control Limit} = \text{Lower spec limit} + (3\sigma_X' - 3\sigma_{\bar{X}})$$

$$\text{or, more simply, LMCL} = \text{Lower spec limit} + (3\bar{R}/d_2 - A_2\bar{R})$$

This relationship and the one following are shown in Figure 2, and the control chart factors are found in Table 1.

Some margin of safety seems necessary when the conventional control limits are replaced by modified control limits. The fundamental principle of the control chart is the establishing of control limits based on the process itself. When these are given up in favor of modified control limits a large part of the useful information given by the control chart is eliminated. For this reason, the author prefers the use of what may be termed 2-sigma modified control limits. The standard formulas are modified as follows:

$$\text{Upper Modified Control Limit} = \text{Upper spec limit} - (3\sigma_X' - 2\sigma_{\bar{X}})$$

$$\text{or UMCL} = \text{Upper spec limit} - (3\bar{R}/d_2 - 2/3\ A_2\bar{R})$$

$$\text{Lower Modified Control Limit} = \text{Lower spec limit} + (3\sigma_X' - 2\sigma_{\bar{X}})$$

$$\text{or LMCL} = \text{Lower spec limit} + (3\bar{R}/d_2 - 2/3\ A_2\bar{R})$$

Pre-Control Limits

Frequently it is desired to construct control chart limits without previous evidence of process capability. To do this, it is assumed that the process spread will just meet specifications, i.e., specification spread equals 6-sigma. From this hypothesis we may derive a value for average range as follows:

$$\bar{R} = \frac{d_2 \times \text{Specification spread}}{6}$$

From this value of $\bar{R}$, the control limits for averages and ranges can be calculated thus: specification mid-point $\pm A_2\bar{R}$ and the appropriate $D_3\bar{R}$ and $D_4\bar{R}$ may also be determined. These may be used until sufficient values of range are secured to make any necessary corrections in the value of $\bar{R}$, $A_2\bar{R}$, $D_3\bar{R}$, and $D_4\bar{R}$.

Efficiency of Averages

The control chart has been defined as an efficient tool for detecting changes in process level and variability. The use of sample averages offers a more efficient means of detecting shift in process level because of a relationship stated earlier in this paper, i.e., $\sigma_{\bar{X}} = \sigma_X'/\sqrt{n}$ which may also be expressed as $\sigma_X' = \sqrt{n}\ \sigma_{\bar{X}}$. This means that a process shift expressed in terms of the standard deviation of individuals becomes increased by the $\sqrt{n}$ when expressed in terms of the standard deviation of averages.

The effect of this is shown in Figure 3, where for averages of samples of four, a shift in process average of $1\sigma_X'$ results in a shift of $2\sigma_{\bar{X}}$ with a resulting ratio of 15.9% of averages out of control compared to 2.3% of individuals out of limit or a sensitivity ratio of about 7/1. With a shift of $1.5\sigma_X'$ this ratio increases to about 7.5/1. This relationship is shown in greater detail in Figure 4 where data for averages of samples of 2, 3, 4, 5, 6, 9 and 10 are given.

An example of the use of this is illustrated by the following problem: After 30 samples of 5 are taken and the values of $\bar{X}$ and R plotted, the $\bar{\bar{X}}$ on a control chart turns out to be .4382" and the $\bar{R}$ = .0121. If the process level shifts to .4315", what is the probability of detecting the shift in the first sample taken after the change actually occurs. The procedure is as follows: (1) $\bar{R}/d_2 = .0121/2.326 = .0052$; (2) $(.4382 - .4315) / .0052 = .0067/.0052 = 1.29$; (3) in Figure 4 find 1.29 on the ordinate; (4) draw a line horizontally until it intersects the curved line for n = 5; (5) draw a line vertically downward from this point until it intersects the abcissa; (6) read the value of probability which in this case is approximately .46. The answer is then that the probability of detecting the shift on the first sample drawn is .46 or about once out of every two samples.

Thus it can be seen that any two of the following three factors are necessary to use Figure 4: (1) sample size (n); (2) process shift expressed in terms of σ_X' ; (3) desired probability of detecting the shift in the first sample drawn after the change takes place.

Typical Control Charts

Since the readers of this paper represent such a wide range of interest and product, the control chart presented in Figure 5 relates to a situation with which the author is sure that all are familiar, the bowling score. This figure portrays the average and range of 3 game scores for 22 nights of bowling on the part of one of the author's associates. The pattern of variation of the averages and ranges shows a random "only to be expected variation", well within the calculated control limits. From this figure, we can determine that the $\bar{\bar{X}}$ is 169 and $\bar{R}/d_2$ is 39.0/1.693 which equals 23. The calculation of $\bar{\bar{X}} \pm 3\ \bar{R}/d_2$ results in a predicted limit for individual games of approximately 238 to 100. This indicates that friend bowler may expect a high game score of 238 and a low game score of 100 if he bowls enough games to realize these extreme predictions. Furthermore, most of the time (i.e. about 2/3, 68.3% to be exact) his score will vary between 192 and 146. This knowledge should make the sport much more pleasant because a low game score (but not below 100) need not be a reason for dejection, neither should a high game score (but not above 238) be a reason for undue elation. Both are to be expected, albeit very infrequently.

Conclusion

Properly set up and interpreted, the Control Chart for Average and Range is a valuable addition to the quality engineer's kit of tools for controlling the process to secure that ever elusive but constantly challenging goal, complete conformance to specification.

References

(1) Shewhart, W.A., "Economic Control of Quality of Manufactured Product", D. Van Nostrand Company, Inc., New York, N. Y.

(2) Simon, Gen. Leslie E., "Engineer's Manual of Statistics", John Wiley & Sons, Inc., New York, N. Y.

(3) Grant, E. L., "Statistical Quality Control", 1946 McGraw-Hill Book Co., Inc., New York, N. Y.

(4) Winterhalter, Alfred J., "Development of Reject Limits for Measurements", Industrial Quality Control, Jan., Mar., 1945.

TABLE #1

FACTORS FOR COMPUTING CONTROL CHART LIMITS

No. of Measurements in Sample	CHART FOR AVERAGES			CHART FOR RANGES		FOR INDIVIDUALS
	Factors for Control Limits			Factors for Control Limits		$\sigma_X' = \bar{R}/d_2$
	Regular	Modified				
n	A_2	A_{22}	A_{23}	D_3	D_4	d_2
2	1.880	1.406	.779	0	3.268	1.128
3	1.023	1.090	.749	0	2.574	1.693
4	.729	.971	.728	0	2.282	2.059
5	.577	.905	.713	0	2.114	2.326
6	.483	.862	.701	0	2.004	2.534
7	.419	.830	.690	.076	1.924	2.704
8	.373	.805	.681	.136	1.864	2.847
9	.337	.785	.673	.184	1.816	2.970
10	.308	.770	.667	.223	1.777	3.078
11	.285	.756	.661	.256	1.744	3.173
12	.266	.744	.655	.284	1.717	3.258
13	.249	.733	.650	.308	1.692	3.336
14	.235	.723	.645	.329	1.671	3.407
15	.223	.715	.641	.348	1.652	3.472

FORMULAS FOR REGULAR CONTROL CHARTS

Chart for	Central Line	3-Sigma Control Limits
Averages	$\bar{\bar{X}}$	$\bar{\bar{X}} \pm A_2\bar{R}$
Ranges	$\bar{R}$	$D_3\bar{R}$ and $D_4\bar{R}$
Estimated spread of individual measurements:		$\bar{\bar{X}} \pm 3\bar{R}/d_2$

FORMULAS FOR MODIFIED CONTROL LIMITS

3-Sigma Control Limits	Upper Spec Limit $- A_{23}\bar{R}$
	Lower Spec Limit $+ A_{23}\bar{R}$
2-Sigma Control Limits	Upper Spec Limit $- A_{22}\bar{R}$
	Lower Spec Limit $+ A_{22}\bar{R}$

A_{23} Factor $= (3\bar{R}/d_2 - A_2\bar{R})$

A_{22} Factor $= (3\bar{R}/d_2 - 2/3\, A_2\bar{R})$

FIGURE 1

RELATIONSHIP BETWEEN INDIVIDUAL MEASUREMENTS AND SAMPLE AVERAGES

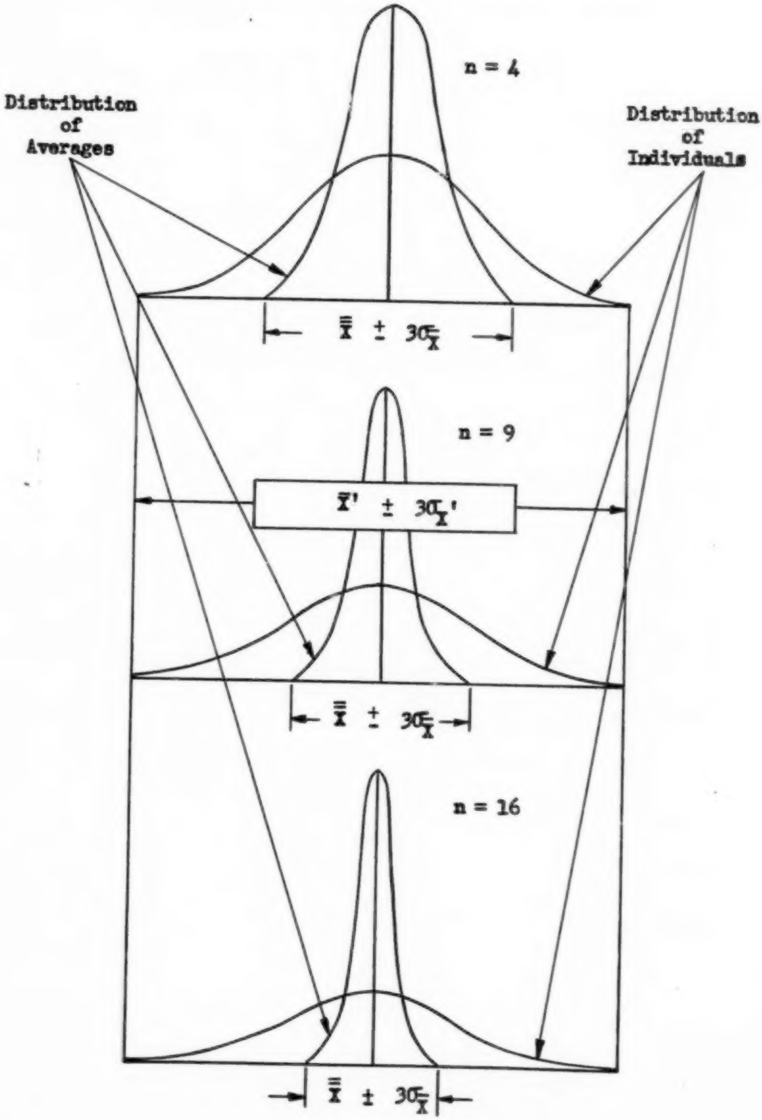


FIGURE 2

THE DEVELOPMENT OF MODIFIED CONTROL LIMITS
for SAMPLE SIZE = 4

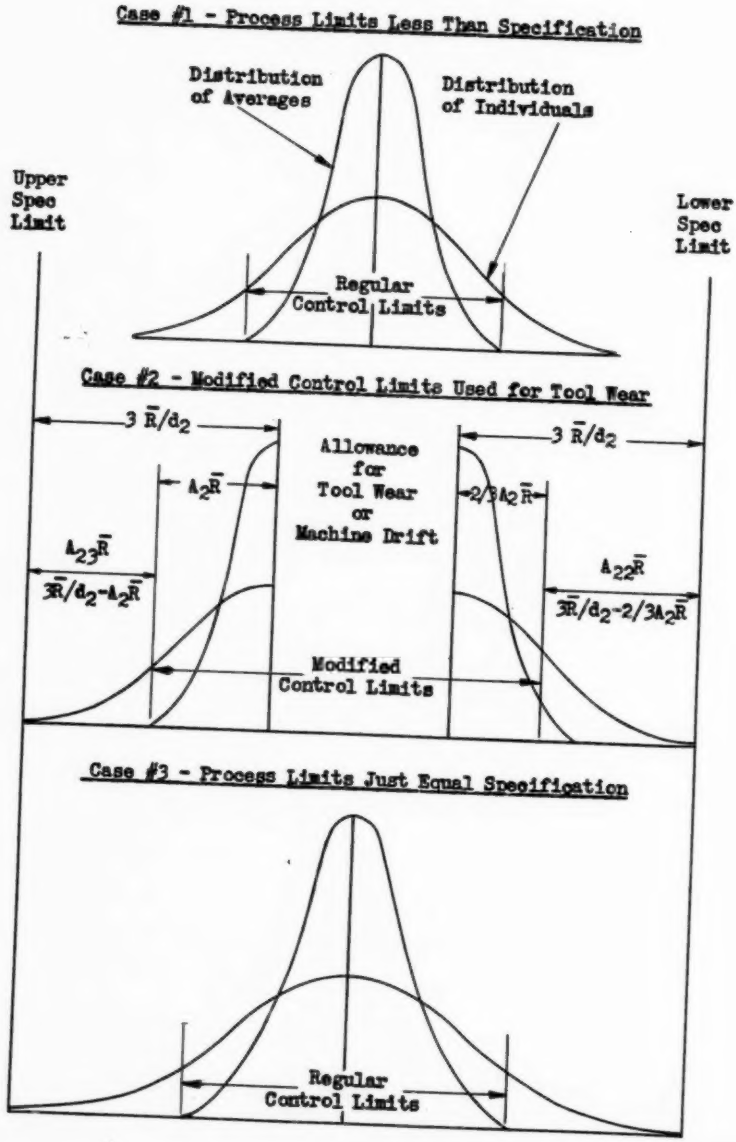


FIGURE 3

USE OF AVERAGES FOR DETECTING PROCESS SHIFT

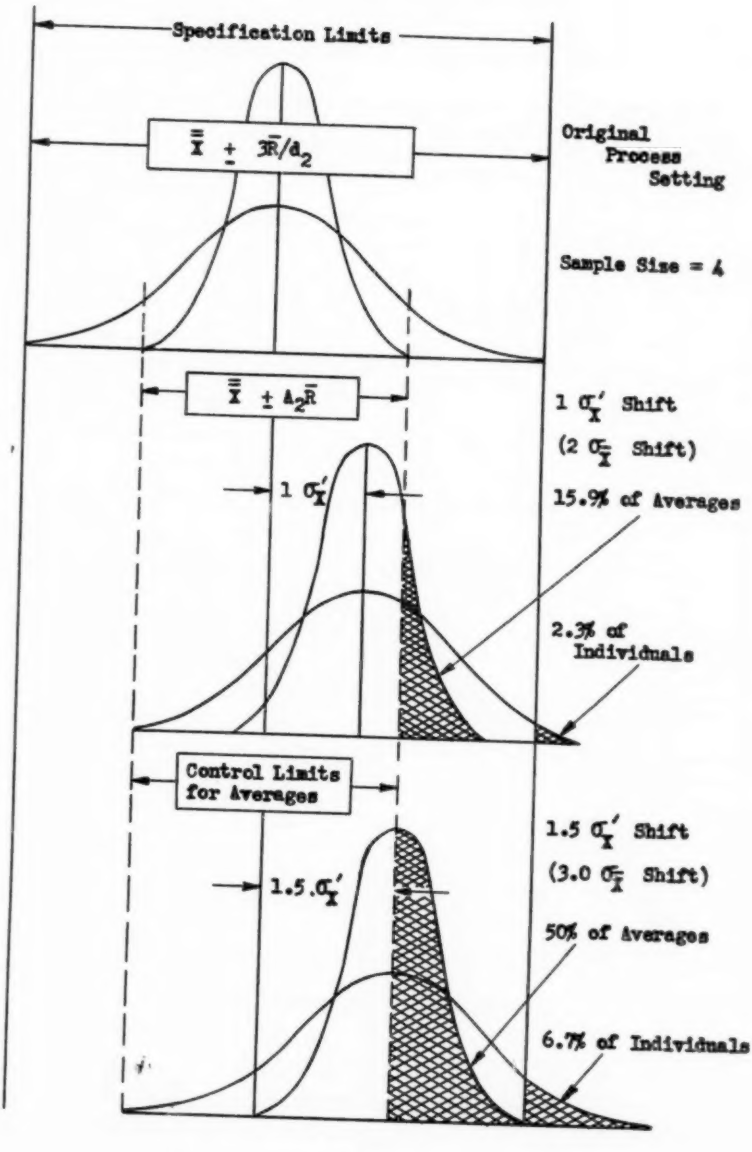


FIGURE 4

EFFICIENCY OF AVERAGES FOR DETECTING PROCESS SHIFT

LIMITS FOR INDIVIDUALS	$\bar{\bar{X}} \pm 3\bar{R}/d_2$
LIMITS FOR SAMPLE AVERAGE	$\bar{\bar{X}} \pm A_2\bar{R}$

NOTE: It is assumed that process is just capable of conforming to $\bar{X} \pm 3\bar{R}/d_2$ or that process is at $3\bar{R}/d_2$ from upper or lower spec. limit.

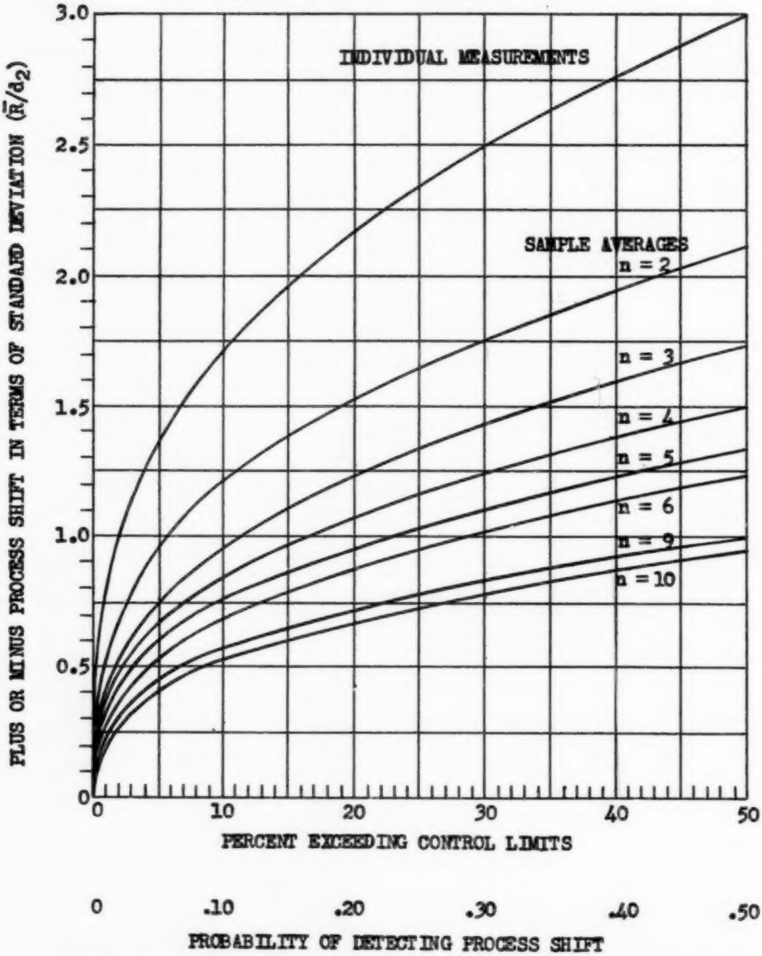
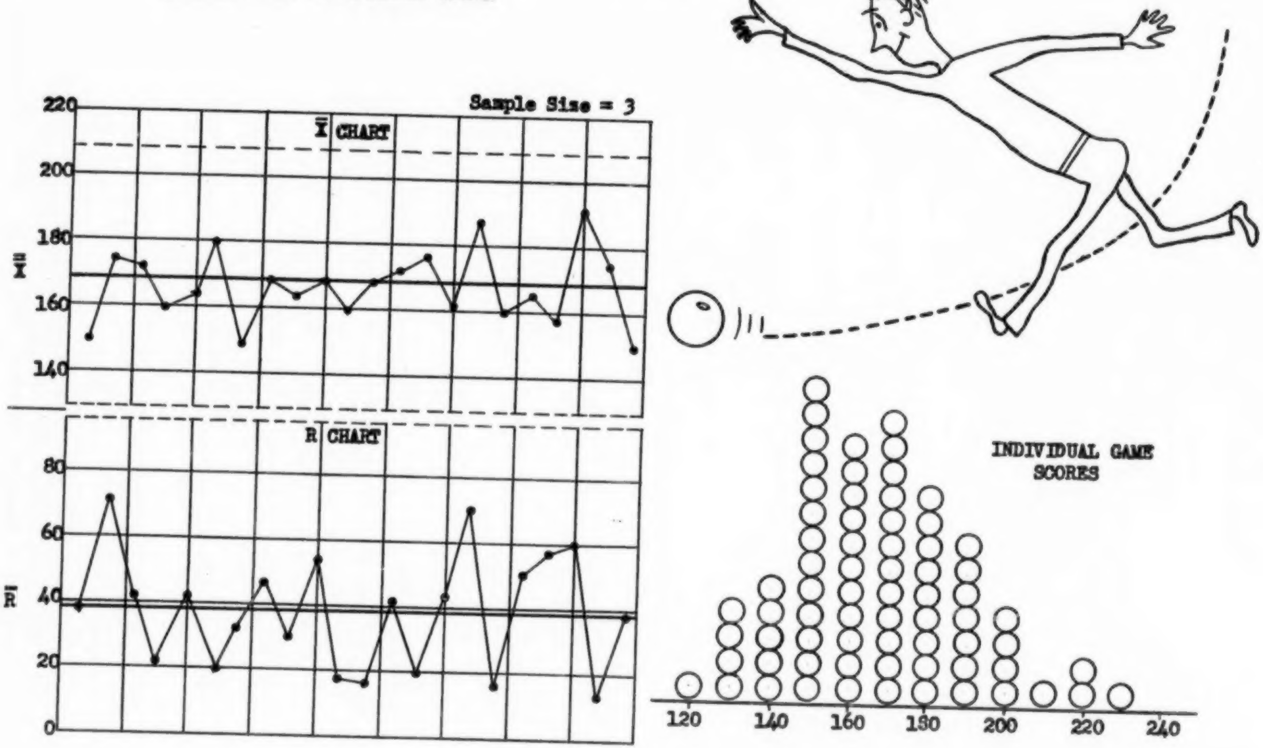


FIGURE 5

CONTROL CHART FOR BOWLING SCORE



A QUALITY CONTROL SYSTEM FOR JOB SHOP ELECTRONIC EQUIPMENT MANUFACTURE

Fred J. Berkenkamp
General Electric Company

One of the basic essentials to profitable and expanding business in job shop manufacture of electronic equipment is a strong and progressive quality control activity.

Through proper organization, effective "feedback," and integrated quality program control, improved quality at minimum quality costs can be achieved despite product variety, technical complexity, very small lot sizes, short manufacturing cycles, and frequent design and model changes.

This presentation deals with:

1. The type-of-manufacture and type-of-business factors which the quality control system must meet.

2. The major responsibilities of the quality control component.

3. The quality control organization and its location within the operating department.

4. The highly essential "feedback" cycle.

5. The work elements of the components within the quality control function.

6. The four major job areas of the quality control activity--new design control, incoming-material control, process control, and special process studies.

7. Highlights of successful procedures and practices in each of these major control areas.

8. Quality control activity performance measuring sticks.

In job shop manufacture of electronic equipment particular emphasis is required in the preproduction quality control activities, and in establishing effective quality "feedback" systems throughout the manufacturing cycle.

Since pilot runs generally are not feasible from a time and dollar aspect, proving-out of new designs before production is generally limited to one or two prototypes. Performance variations due to circuitry and component variability cannot be accurately gauged from one or two prototypes. As a result, the design engineer must include an economic safety factor in his design to cover this contingency.

In manufacture, the entire production lot may be largely assembled and wired before tests are completed on the first production equipment. Serious performance deficiencies found at this time can be highly expensive in terms of correcting the remainder of the production order. Unless heavy emphasis is placed upon preserving this design safety factor for normal circuitry performance variations, profits can be quickly absorbed by scrap, rework, and excessive inspection and test costs.

It is essential that this design safety factor not be used up through excessive manufacturing tolerance buildups, inadequately planned machine shop and assembly methods and practices, lack of adequate attention to critical parts, wiring, dimensions, finishes, etc.

To provide the essential pre-planning to meet this goal, positive preproduction quality control effort must be applied with the Sales, Design Engineering, Manufacturing Engineering, and Purchasing groups.

Small lots, short manufacturing cycles, high product variety, high per unit value, technical complexity, frequent design and model changes all place heavy demands on the "feedback" systems in the manufacturing cycle from Sales, Engineering, etc., through Shipping.

The information fed back must be current, relative and effectively presented to generate positive action. Through properly designed record keeping and tabulating systems, rapid, accurate analysis of a wide variety of data is easily obtained. A number of feedback systems meeting these demands are discussed.

Finally, quality costs and quality audits, the basic performance measurement tools of the quality activity, are reviewed. Emphasis is placed upon interpretation and evaluation of these tools for corrective action in obtaining optimum quality at minimum quality costs.

EVALUATION OF RELIABILITY IN GUIDED MISSILE SYSTEMS

Gerald R. Sams
U. S. Naval Ordnance Laboratory
Corona, California

Evaluation and reliability methods and procedures have been a popular subject in recent years, particularly in the guided missile programs. This paper will deal with methods of determination and control of reliability.

It goes without saying that classified information will be excluded from this paper.

History has shown that during the early stages of evolution most complex devices have been considerably unreliable - for instance the automobile or radio.

It is not difficult to understand the importance of reliability in guided missiles. Economics, logistics and strategy demand high reliability. Because of these factors there is considerable interest at high Department of Defense levels in the reliability of guided missiles.

Evaluation can be defined as a process to determine the ability of a product to perform a given function.

Reliability can be defined as the trustworthiness of a product to perform a given function.

Evaluation could be defined as a process to determine product reliability. I favor this latter definition and will present in detail a guided missile reliability program.

There are at least two methods that can be used in the evaluation of complex devices such as guided missiles. One method involves a very formal and thorough test program after the missile has been designed, developed and produced. The other method requires an evaluation on an incremental basis as the missile program progresses. The first method is very effective if properly implemented. If the designer, developer, producer knows that his product is going to be subjected to a rigorous test and evaluation program he is more likely to pay attention to its more subtle and hard-to-discover faults. Quite often the pressure of time and the shortage of manpower tempts him to take short-cuts, or calculated risks without adequate information with which to make the calculation. This can, and has, led to serious problems that almost defy logical solution. I believe that most of us here understand the formal-and-thorough testing method of evaluation so I will describe in detail the incremental method of evaluation.

It is my opinion that an evaluator must have two characteristics - i.e., objectivity and subject matter competence. The designer-developer has the subject matter competence but it is questionable that he can be completely objective about his creation. On the other hand an independent evaluator may be completely objective but less than completely competent technically. If one follows this line of reasoning it is readily apparent that the designer-developer should not perform the formal evaluation. Nor should anyone else with a vested interest perform it.

The method I am about to describe represents a sensible program without tricks or magic. In fact, this presentation is important only because the method is so rarely followed with diligence in guided missile programs.

As I have said before, the pressure of time and shortage of qualified manpower tempts one to take calculated risks, based on engineering intuition rather than known facts.

The evolution of nearly every weapon system can be divided into fairly distinct phases - design and development, product engineering, production and storage, and use. Although these phases may overlap somewhat in time, each phase is quite distinquishable one from another. It appears desirable to organize this paper according to the several phases of evolution. Action should include control measures designed to hasten the transition of a reliable system from one phase to the next. This should include the best possible estimates of the current reliability of the system based on available data. To permit continuity and proper control of reliability through the various phases, careful and complete planning for future reliability measures is required.

DESIGN AND DEVELOPMENT PHASE

When concerned with the collection and analysis of data for reliability control, confusion of terminology can be a serious handicap. It is in order to fix a language for description of the missile system. Having fixed this language the system should be divided into its major assemblies (units) and each unit into sub-units down to the part level. When properly defined the units of the system could serve as units of development, units of inspection, and units in the chain for estimation of system reliability. Physical size and complexity of the unit requires careful consideration. If the unit is to serve as a basis for life and environmental testing it must be small enough to permit economic testing of adequate samples and large enough to restrict the total number of tests required to a reasonable size. Such units, whose reliabilities are to be individually monitored, should have a characteristic transfer-function which can be measured and monitored both when the unit is assembled in the system and when it is treated as a separate entity. The unit should be of great enough importance to the system that reasonable assurance of successful flight is given whenever all unit functions are determined to be within specified tolerances.

The first reliability consideration in the evolution of a missile system should be a careful review of design proposals for the system. The purpose of such a review should be, primarily, to determine whether the same or similar functions could be performed by a simpler system or even by several simpler systems having proven high reliability.

It is necessary, at the start of this phase, to obtain some estimate of the environmental extremes in which the system will have to live and operate. These estimates would serve at least two important functions:

a. Provide criteria for design.

b. Provide preliminary limits for planning environmental and life tests.

During the design, the system proposed must be evaluated to determine whether the basic ideas, as expressed in the design and design requirements, can be realized. Careful thought and planning should be given to the need for documentation of test procedures and standardization of test equipment in order that development can be effected without confusion or delay in the collection of useful data for analysis.

Later in the program it will be necessary to have the several agencies which will be performing tests on elements of the system, for the same or similar purposes, use, as nearly as possible, the same procedures and equipments. Otherwise, a great deal of validity will be lost in conclusions drawn from their combined results and in many cases separate groups will arrive at different conclusions about the same item due solely to differences in procedures or equipments.

An essential part of a reliability program is the preparation, in advance of development, of an integrated plan for the collection, analysis and dissemination of data among all contractors and cognizant agencies in the program.

The need for data from tests to be in such form that it can be correlated among the several sources cannot be overemphasized.

A further source of useful data is the failure reporting system. This too should be coordinated among all groups likely to observe failures of the system or any of its parts. Such coordination should include: a common definition of what constitutes a failure; a single form for recording the failures; and a single, simple and comprehensive set of instructions to be followed by all engineers and technicians who will be completing the forms. The desired method of summarization, analysis and reporting should be determined, as nearly as possible, in advance; so that as soon as failures begin to occur, periodic reports for timely use by all interested agencies can be prepared in a routine fashion. It is pointed out that for failure data to serve a really useful purpose, it must be supplemented by knowledge of how much opportunity to fail had been given to the faulty units.

The missile log system, carefully planned, can be another source of data.

A great deal of the planning needed for a quality control program should begin during the design phase. For example, workmanship inspection and receiving inspection, with the necessary record formats and charts, etc., should be ready for implementation as soon as possible after a contract is let for model shop fabrication of development prototype models. Only if defects due to workmanship are eliminated or accounted for can a valid evaluation of design and engineering be made from such models. A supplementary yield of workmanship inspection should be records of areas which in the future could cause production delays or difficulties.

DEVELOPMENT

During development the missile system is subject to many changes. Changes are introduced to produce better performance or equivalent performance with greater reliability. In either case the ultimate objective is the elimination of the principal causes of unreliability. An adequate system of failure reporting can serve to point up the major areas of

difficulty and thus contribute to improvement in reliability. However, the actual evaluation of reliability of the system and its individual elements must be based on determination of the ability of the system to perform satisfactorily under the conditions in which it must operate tactically. Such performance includes survival of the hazards of transportation, handling, and storage as well as operation in severe flight environments.

Economic accomplishment of a reliability program entails an adequate environmental testing program for the parts of the system separately, when required, as well as for the system as a whole. This, in turn, requires thorough and accurate knowledge of the environmental hazards to be encountered and simulated.

Test programs to measure these conditions should be given a high priority as soon as the feasibility of the system has been demonstrated.

Environmental and life tests are fundamental for orderly, economic and continuous evaluation of the system. If the tests are designed with extreme care to assure both engineering and statistical adequacy, the results, properly interpreted, are of great value for evaluating progress.

Frequently tests to failure should be included to provide a measure of "strength" where samples are statistically inadequate.

During development nearly every missile produced is in several ways different from its predecessors. Most models are handmade by highly trained engineers and technicians. Each man concerned in the fabrication of a model is, by his training and experience, competent to devise appropriate functional tests to check or adjust the units and systems after each step in the assembly of the model. These procedures are seldom written, and unfortunately, two equally competent men are not likely to devise procedures which are identical in process or precision. Since the tests and adjustments are often duplicated by different men, and frequently at separate activities, it is desirable, from the standpoint of getting useful and comparable data for analysis, to document and standardize such test procedures as early as possible.

The documentation of procedures for review by others performing similar tests allows comparison of techniques so that by the time the procedures are needed by factory production personnel they will be ready in nearly optimum form.

Similar arguments hold for inspection procedures which should be instituted and checked as early in the program as possible to permit a smooth transition from model shop to factory production.

Documentation and maintenance of test and inspection procedures may require full time attention of specialized personnel. The persons involved should determine by observation and study, and with adequate and competent technical guidance, the fabrication and testing techniques and procedures, as well as anything else which could influence design of test and inspection procedures. These, when written into a standard prescribed format, then form the basis for the design and documentation of a complete set of standard procedures.

Although the procedures will be followed initially by the more

skilled technicians of the model shop they should be prepared in complete detail, step by step, so that they are intelligible and simple to understand and follow by the less skilled personnel of the production assembly line. Accompanying data sheets should be simple but complete and comprehensive as required for engineering and statistical analysis. They should be so organized that each succeeding recording corresponds, in sequence, with the succession of steps in the procedure.

Statistical quality control is generally considered a tool for mass-production. However, in addition to being a collection of useful specialized techniques for determining quality levels of production and assuring a desired quality level over long production runs, quality control includes systematic and logical methods for collecting useful data from inspection and production tests, etc. In this sense quality control is useful even in "small lot" production, and the basic records which will later be useful in production are useful also during model shop fabrication, and their employment should be instituted as early as possible.

PLANNING FOR NEXT PHASE

When the development of the system has progressed to a stage where the missile can be considered seriously as a mass-producible weapon system, the program must have available a procedure for final inspection or system "check-out" to permit reliable determination of the quality of individual missiles. Such procedures would be applied to assembled guidance and control systems to compare system performance to design specifications.

The principal requirements for such procedures are that they should: (1) be accurate, so that acceptance or rejection can be easily established in a high proportion of cases; (2) be rapid, so that the final inspection workload is not a bottleneck; and (3) have no detrimental effect on the system, so that the predicted future performance will not be changed adversely by the test.

In order to assure accuracy and validity the preparation of final check-out procedures should begin with a comprehensive analysis of the missile system to determine which parameters can be used most validly, to predict performance.

When a system parameter is a function of several sub-assembly parameters, only the system parameter should be observed, if feasible, since the sub-assemblies will, presumably, have been fabricated properly, inspected, calibrated and tested.

The check-out procedure and the necessary equipment should be designed to assure adequate performance when tests are conducted by technicians because of the unavailability of engineering talent during national emergencies.

Finally, the test must proceed rapidly so that the useful life of the missile is not reduced.

Consideration should also be given to methods of abbreviating and further mechanizing the test for use as a tactical periodic check-out. The space required for such check-outs should be minimized. The test should be of a simple go-no-go nature with procedures which are simple and automatic enough to permit performance by relatively unskilled

combat personnel.

An important factor in the reliability of a missile system is the human element. Since the human effect on reliability will not, generally, be determined, the only recourse is a comprehensive training program for the personnel involved so that the chance of human error is minimized.

The effect of training will be felt equally in the reliability and evaluation program as well as in the reliability of the system. The reliability program is dependent for its success on the quality of the data available. Thus, the training program for all who will be responsible for generating and recording data, must place strong emphasis on the importance of precise and accurate measurement and complete recording of all the data requested.

The primary objective in the design of a shipping container for a missile or its components is to provide them with protection from those hazards of transportation, storage, and handling which are likely to have an adverse effect on the reliability of the missile.

In order to evaluate the protection afforded by a container information must be available on:

a. Assessment of the types of hazards to be encountered.

b. Ability of the missile (or component) itself to withstand adverse conditions.

c. The effectiveness of the container to provide protection for the missile or components against those hazards which the missile (or component) alone cannot withstand.

d. Value of container in relation to the value of missile damage avoided.

It is economically and logistically unfeasible to protect the missile against all possible contingencies. Protection against the shock resultant from a train wreck, for example, is impractical; the probability of occurrence is extremely small.

When the extent of hazards and the capability of the missile to withstand such hazards are known, the criteria for evaluation of the container are available. The container must be such that the container-missile combination can withstand the total environment of transportation, storage, and handling with high probability. It should not, however, be better and thus generally more expensive, than is required.

Prior to the beginning of a large scale production program there should be a thorough investigation into the capability of every unit of the system to operate for the required length of time in its flight environment, after being subjected to the environments of fabrication and production testing, transportation, handling, and storage. Such a test program is often appropriately called a Type Approval Test (TAT).

Unfortunately, in many cases, a unit type is accepted on the basis of a TAT on a single item. The difficulties inherent in evaluating guided missiles dictate the use of adequate samples for the type approval test program.

In addition to environmental tests the TAT program should include bench life tests premised on the best available estimates of the total operating time required to test, adjust and check the missile and its units from assembly through surveillance and periodic check-out to launching.

After a unit is subjected to a simulated and perhaps accelerated life it should be tested to destruction in severe environments to determine its "breaking strength" and its capability of operating in severe environments for the required flight time. Two conditions are necessary to the adequacy of this phase of the TAT program:

a. Environmental tests must be conducted in combinations of severe environments, not, as is often the case, sequentially in each separate environment.

b. The environmental severities should be sufficiently great to provide a safety margin against the factors which are ignored or unknown.

Only after a unit has passed its TAT should its design be released for mass-production.

Reliability and producibility are both important military characteristics of the system. Thus, in some cases, one characteristic must yield in a compromise to improve the other. Every such compromise should be arrived at only after a comprehensive review of the effect on the whole system, not, as is so often the case, after consideration of the effect on the particular circuit or assembly involved.

Before a realistic evaluation of engineering adequacy can be accomplished, the variations in performance and the proportion of failures due to material and workmanship defects must be determined. There have been many cases in the past of complete inability to determine whether a series of failures of a unit were due to inadequate design or poor workmanship because the records of inspection, kept in the shop, were poor if indeed they existed at all.

When failure reports are supplemented by data yielding a measure of "opportunity to fail", the resulting calculated failure rates are an excellent first measure of design adequacy. That portion of the failure rate attributable to poor workmanship can be regarded as a limiting measure of producibility of the design, since the fabricators in the model shop can be presumed to be highly skilled technicians, rather than unskilled labor that may operate a production line. The remainder of the failure rate, not attributable to material or workmanship defects, serves as a measure of design adequacy in the sense that it estimates the probability of an "initial failure" of the unit. Performance data from successful tests should yield estimates of the mean and variations of unit parameters which, when compared to design tolerances, also serve as measures of the ability to produce satisfactory units to design specifications. If a disproportionate number of units fail to meet such specifications the tolerances must be reviewed to determine the feasibility of widening them and, simultaneously, the design and the assembly process must be reviewed to determine the feasibility of increasing producibility and "tightening" production.

Simulation programs are a useful source of data to determine:

a. the effect of variation in unit performance, in particular of marginal performance, on flight characteristics,

b. the effect of design changes, and

c. the susceptibility of the system to countermeasures. Simulation programs, when realistically conducted and carefully correlated to comparable flight results, permit observation of statistically adequate samples with comparatively small expenditures.

As a result of the ground test program and the evaluation flight test program the following information should now be available:

a. Probability of no "initial" unit failures due to defects.

b. Probability that all units operate for the required time given no initial failure.

c. Product of a and b equals inherent flight reliability.

d. Correlation between missile periodic check-out and flight reliability (a measure of the adequacy of check-out procedures and equipment).

e. An upper bound on tactical kill probability given by an appropriately weighted average of the proportion of successful flight resulting in "kills".

PLANNING

In a situation where the goal is production of large quantities of high quality articles it is necessary that these articles be all of a kind. It follows that the barest minimum of minor changes can be permitted to disrupt the pattern of repetitition. In cases where the design of the article to be produced is not, or cannot be, frozen mass-production methods must be compromised.

Design changes and modifications may be introduced in block form, i.e., they may accumulate to be introduced into production of the kth and all subsequent items. The optimum block size will usually be determined by compromise among such considerations as the need for utilization of mass-production methods, the ability to introduce design changes as soon as their desirability is well demonstrated; the need for missiles for training and other purposes, the skill of the workers available, etc.

Modifications must be tightly controlled from the view point of reliability to the extent that they must be of proven value, in terms of either improved performance or increased economy, prior to disrupting production by the introduction of the change.

When the reliability level which is achievable by the production design has been determined, an important step in the reliability control program is one designed to keep constant check on the quality level of the assembled missiles. A proof test firing program, which calls for firing

of small randomly chosen samples from production lots, is one valuable means of providing this information.

The principal objectives of a proof test firing program are (a) to demonstrate that the quality level established earlier is being maintained or improved in production; (b) to provide, on a sampling basis, an estimate of the quality of a production lot; and (c) to provide additional engineering data for feedback into design and production groups.

PRODUCTION AND OPERATIONAL PHASE

Even during production and operational use it is desirable that tests of units of the system in environmental conditions simulating future experience, should continue. These tests should be standardized and may be employed as part of the quality control program to catch production lots of weak units before they cause rejection of entire missiles. Such tests are generally called "type tests" and serve a role for "units" similar to the role of production proof flights for the complete missile.

In addition, as more and more missiles are flown for training and during simulated battle maneuvers, the continued analysis of flight reports and flight failure reports may point out new areas of weakness which were not apparent during flights made under test conditions to obtain engineering data. Such cases, should bring out a new cycle of design, (to eliminate the weakness) and tests (to demonstrate that the weakness has been removed).

Changes in the expected environment of transportation and flight, etc., should result in changes in the environments in which units are tested to assure the production of future units which are capable of surviving in those environments. The reliability program should continue to monitor the quality control activities of the prime contractor to provide assurance to the sponsoring organization that quality requirements will be met even though inspection personnel representing the sponsor is kept at a minimum. Results of failure analysis and flight reports should be fed into the reliability program to ensure the use of sampling inspection schemes that provide the protection required by the sponsor.

The proof test firing program, and the analysis of the results should be directed and reviewed by an unbiased group.

It is not sufficient to design and produce a high quality missile and to package it carefully to protect it against shocks in handling and transport and against meterorological extremes of temperature and humidity in storage. If the missile is to be kept "on the shelf" for any period, and then be used with any degree of confidence, the deterioration in its quality, resulting from storage must be known or predictable. This measurement of the rate of deterioration is one of the primary objectives of a storage surveillance program.

Where there are good estimates of shelf life of the missile system and its parts, obtained in an adequate variety of storage conditions, a rework and replacement schedule can be prepared which will call for automatic replacement of potential sources of system failure at a time so chosen as to give a high degree of confidence that all such units, that

may have deteriorated beyond necessary limits, have been replaced prior to use. Since "freezing" the design of a missile system for production seldom means the end of development and improvement of the system, the various flight programs which begin, or continue, with the beginning of production should be a principal source of data for engineering and evaluation.

In most cases, effort to assure collection of correlatable data will not add greatly to the cost and effort required for the test. Such data would be extremely valuable for assuring a continuous reaffirmation of the reliability of the system and of the validity of the conclusion that the system represents a usable and useful service weapon.

MARKET RESEARCH SETS QUALITY CONTROL TARGETS

Theodore H. Brown
Harvard University
Graduate School of Business Administration

Some years ago and just two weeks before Christmas one of the top-flight department stores here in New York opened five thousand dozen pairs of women's hose in preparation for the last-minute Christmas shopping. To the buyer's dismay, it was found that every pair was defective. Regardless of circumstances, customer demands had to be met. The buyer consequently was forced to purchase in the open market an equivalent amount of good merchandise at whatever price was asked.

This costly experience raises three major points at which quality could have been checked. First there was the need for quality control at the manufacturer's level. With this you are all familiar. Second there was no acceptance sampling for quality on the part of the department store buyer who failed to make even a spot check. He believed that this was not necessary because by the custom of the trade any defectives could have been returned to the hosiery mill in exchange for perfect merchandise. Third there was the supreme court of last resort at the customer level. Appeal to this court was not made, since everyone knows that Mrs. America demands in the purchase of her quality hosiery practically a 100% perfection.

It is unnecessary to use market research in order to find out whether Mrs. America wants perfect hose. Every one of you gentlemen is ever mindful of crooked seams, wrong color, rings, runs, and tubular hose without style. There are, however, many problems such as quality of finish and shades of color which are usually called elements of fashion, but which the creators of fashion and the buyers of hosiery must know if they are to be successful in keeping up with the market. Market research is designed to resolve these questions and so to secure more nearly objective answers than is possible through an emotional appraisal.

The question then is just what information does market research provide toward the solution of problems in which qualities of products are important. For this purpose we select from a wide range of objectives three which apply here.

First the products of the manufacturer and merchant must possess those qualities which make them more saleable in the market. If this is not done, some smart competitor will incorporate these same qualities in his own lines of merchandise to increase his own sales. Consumer wants are never satisfied. In addition the imagination of people in the American market creates demands for definite qualities in products which make them more desirable. It is this continual search for more saleable products which keeps businesses alive. Two cases will illustrate this question of saleability.

Immediately following World War II, the engineers in a subsidiary of a well-known company proposed that the parent company engage in the manufacture and sale of a tape recorder to be used by executives as a

dictating device from which secretaries were to transcribe letters and memoranda. The product looked good because the design qualities of the models seemed to be excellent. Nevertheless, a field test among a number of secretaries in a large city showed that the model had many shortcomings. One of these was the difficulty in locating a particular piece of dictation on the long tape. Other similar unsuspected difficulties turned up in this market testing. The recorder as it had been built in model form was not saleable in the competitive market. The parent company consequently dropped the project which might have brought them nearly a half-million dollar failure.

The question is not always an open and shut one. Sometimes a product will be accepted by one part of the market but rejected with scorn by another. A personal experience not long ago again brought this sharply to mind.

Late one evening I stopped for a light meal of oyster stew at a seafood restaurant in an eastern seaboard city. On ordering, I requested no pepper because I happen to enjoy thoroughly the taste of oysters. To my horror I found that there must have been included a double dose of Tabasco sauce with plenty of black pepper, but following my request, there was omitted only the paprika on top. It took two bars of chocolate to smooth up the portal to the final resting place where all good food should go. On further inquiry from the waiter, I discovered that currently he thought that the city residents like the product made that way. But do they? Who knows? What is the saleable product? Is it Tabasco sauce taken straight or should it be diluted a little with oysters and milk to bring out its real quality? I doubt whether the brand of hellfire and damnation served me would be saleable in Providence (R.I.). The quality was there, but was it of a kind which makes oyster stew really saleable? -- Only market research would tell.

The second of our market research problems is that of appraising the preference of consumers for one product as compared with substitute or competing goods. For many cases the question here is to find those qualities which make a product wantable. This is one of Shewhart's classes of qualities. Some illustrations again may help.

In one market research project consumers were asked to appraise the relative qualities of competitive lines of consumer brand merchandise. For example, they were asked whether they had ever heard of each of four or five different brands of toothpaste. Then they were asked to rate each brand in terms of its quality as good, average, or poor. Presumably these consumers knew nothing of the chemical quality of each brand. Their opinion was based on a subjective emotional belief concerning the qualities present.

A second illustration is that of the acceptability of a machine to manufacture commutators for fractional horsepower motors. Prior to the war, commutators were made in several steps which may be summarized as manufacturing the parts and then assembling by hand. Since many commutators were needed during the war for the thousands of motors used in airplanes of the Air Force and other military services, an automatic machine was developed for building these commutators. After the war, both a change in the size of the market and design improvements of

motors reduced the potential. Aside from these questions there remained the problem of whether manufacturers of consumer goods would be willing to change their commutator drawings so as to specify standard sizes and whether they would be willing to give up hand assembly even though mass produced commutators were equal or superior in quality. These were the key questions which made the product wantable. If manufacturers should agree to these questions, they would be really specifying certain qualities of the finished product and co-ordinately the qualities of tools to build the product.

Furthermore, this problem of wantable qualities involving market acceptance of a product can be illustrated from everyday life.

In a small New England restaurant a family was having a light noon meal. The man of the family spied some doughnuts on the counter which appeared to be unusually attractive. The lady of the party deciding that the doughnuts were not fit food for that particular man and meal, tried to argue him out of his market demand. Finally when the waitress could stand it no longer, she commented, "Lady, if he wants doughnuts, let him have doughnuts." History does not record whether he got his doughnuts or whether the lady would have agreed if the doughnuts had been cinnamon, chocolate, or the fried doughnut holes which some of us prefer.

I need not extend the picture to the situation where man and wife buy a new icebox, or the family buys a new car, or junior knows just what he wants in a TV set. Decisions in all of these cases prescribe qualities which in themselves set quality control targets or determine the conditions for such targets.

Finally market research is used for guidance in the design and in the manufacture of a product which will possess more desirable qualities marketwise than it otherwise would have had. A very interesting case is that of the development of the moving picture known as "The Jolson Story." Here the market research was carried on co-ordinately with the development of the picture. Some of the high spots in the making of this picture were as follows:

Originally the script was classed as a B or second-grade picture. Presumably it would gross not over a few hundred thousand dollars in revenue. One of the first steps was to find the title most attractive to the movie audience. In succession, the market was tested for the following titles among others: "Minstrel Boy"; "The Story of Al Jolson"; and "The Jolson Story". I believe that everyone will agree that the last was the best. Among other subsequent tests was that of the subject matter, the songs which were proposed, and the popularity of the cast. Again after the film had been taken, a preview was held with a specially selected test audience. In this test, the audience had an opportunity through a mechanical gadget to register their combined opinion from scene to scene. The composite from this test was a curve of opinion whose length corresponded to the length of the film and whose height at any scene represented the test audience opinion. Extreme low spots indicated bad scenes requiring revision, cutting or retaking. Finally the advertising and promotion of the finished film was so timed that the "want to see" build-up was reaching a maximum when the picture was released for public showing. The result was a fine popular picture which

grossed well over a million dollars. In fact, its qualities were regarded so highly that the studio produced a second film under the title, "Jolson Sings Again."

The claim may be promptly made that this last case is one for the creators of art and not one for the quality control engineer whose responsibilities are to be found in maintaining the technical quality of a manufactured product. Nevertheless, it would seem that this illustration indicates the importance of studying the market to discover those qualities which in terms of our first marketing objective will make it more saleable. Building wanted qualities into "The Jolson Story" fundamentally is no different from the work of building desirable qualities into a durable piece of machinery. Thus everyone can understand the wrath of a former student who had the following experience.

Right after World War II a young man purchased in Los Angeles a new high -priced automobile. On the way East, the first thing that happened was that one side of the front bumper dropped off. A day or so later the whole dashboard came loose and hung by a few wires and a speedometer cable. To cap the climax, while going through Harvard Square in Cambridge, Mass. at a speed of 10 or 15 miles an hour, the camshaft snapped. The qualities he desired were not there. The sequel you know. The manufacturer of this particular car and his competitors have introduced quality control so that gross blunders have been eliminated and many new qualities added which purchasers want. General Ayres once remarked that the automobile industry survived because it kept buyers in a perpetual state of discontent. -- It builds into new models new wantable qualities.

The central thought of this paper is that market research will define those qualities of merchandise for which quality control is a must. In a buyer's market the ultimate consumer is the one who makes the final decisions. The decision in some instances may depend upon the presence or absence of specific qualities. These have been called the properties which make a product saleable. The tape recorder and the oyster stew illustrated the "take it or leave it" buyer's choice. Then there is the more complicated situation where the buyer may decide between several products equally good in a technical sense. Preference here is dependent upon some subjective, emotional or hearsay evaluation of the article. We illustrated this problem by the choice of one of several brands of toothpaste by the attitude of manufacturers toward the possibility of standardized commutators, and on the human side, by the emotional belief of whether a doughnut was a suitable product for a man to eat. Finally there was the more difficult problem of building into the final product those qualities which are desired by the maximum number of buyers. "The Jolson Story" illustrated this. The high-priced automobile was used to indicate that in a post-war seller's market a poor product might receive temporary tacit acceptance, but would not be acceptable in the long run.

The quality control engineer may naturally ask at this point, how is all of this to be discovered? In a single sentence which covers a multitude of difficulties, the answer is, "By inquiry from a sample of ultimate consumers." This raises the question of how the sample is obtained and what are some of the problems that the market research worker has to face.

Imagine a quality control engineer in charge of an automatic screw machine division of a super-colossal company operating 110,000 machines - yes, 110 million. This number is not too far different from the number of U. S. individuals who exercise buying decisions. Even worse, imagine that these machines are of all ages, different makes, different capabilities, and are operated with raw stocks of different makes, different capabilities, and are operated with raw stocks of different qualities. Because there are so many machines, a single shop one-half mile wide starting just west of our eastern mountains and stretching far beyond Chicago is to be imagined as broken up into many smaller units scattered over thousands of square miles of the United States. Suppose now that all of these shops are turning cut a single product which can be described only in general terms but that even in a single shop the product is not necessarily uniform nor does it follow any given blueprint. Finally, in each shop some machines are likely to be idle. These may be the best or the poorest machines - just which is unknown.

The problem is to describe the qualities or products which, on the average, are being produced by this crazy pattern of shops. Instinctively your comment is that the first job should be to get some system out of the chaos present. Nevertheless, if each machine is replaced by an individual who possesses his own characteristic ways of doing things, and who has individual likes and dislikes, we have a mass of attitudes which have to be appraised. This is the job the market researcher faces.

One trouble with the situation just sketched is that things are happening without any particular sense of direction. In the super-colossal shop of milling machines, it would help very much if there were a blueprint of the product to be made. This is obvious. Correspondingly in market research a precise blueprint of the objective is of primary importance even though the executives may be in a rush to get on with what they think is the production part of the job. Shop men have their troubles here too. Actually any failure to identify completely and precisely the objectives is the first spot where market research may go wrong.

Maybe the real question of the seafood restaurant is not whether people like stews Tabasco hot, but rather who goes to this restaurant and why? Some go for raw oysters and raw clams; some are very probably travelers who stop because they have heard of this restaurant but who after an experience like mine say, "Never again." Maybe these are not the right objectives but rather our research should ask about the training of the cooks. Other possibilities will come to mind, but one or more must be decided upon as the objective before the investigation starts.

When the definition of the objective has been fixed, a sample may be designed to reach that objective. The basic condition is that each unit of the population from which the sample is to be drawn shall have a chance of selection equal to every other unit. For the 110 million buyers in the U. S. it would be impractical to write every name on a slip of paper and then draw 500 or 1000 for the sample. Moreover, a sample drawn in that way might come from a single large city like Chicago. To avoid this trouble, the 110 million cases are "stratified." A simple way of doing this which is illustrative is to sort out the 110 million people according to the size of the town or city in which each

lives. Then one or more towns in each size class may be picked by chance to represent that whole class size of towns or stratum. In turn within each town so chosen smaller areas are chosen by lot which when put together will represent the town and finally individuals within the small areas are identified by lot.

Assuming that if the stratification and chance selection of towns, areas and finally individuals is done correctly, the summary of data obtained from such a sample will reflect the whole population. The problem obviously is a complex one which requires many hours to complete.

More than once I have heard quality control engineers remark about the difficulty of supervising properly their inspectors scattered throughout large shops. Our 110 million screw machines scattered in small shops scattered over thousands of square miles and more realistically the 110 million potential buyers scattered across the land in towns and villages presents a very difficult problem of supervision to the market researcher. One consequence of the stratification of a sample design which has just been described is that the interviewers can be located at definite points where a group of people who are to be questioned are located. Clearly this saves what would otherwise be a terrific amount of expense in travel for the interviewer. Nevertheless, the central office must retain good supervision even though the interviewers are scattered at definite sampling points from Los Angeles to Boston and from Minneapolis to New Orleans. The obvious way to secure such supervision is through the use of a few traveling supervisors.

This question of supervision of the work of interviewers is more difficult than the corresponding work in the shop. The questions to be asked in market research are not the simple ones such as, "What is the micrometer measurement of this or that dimension?" They are complex questions biased by the common idiosyncrasies, desires, and prejudices of living people. Hence the training of interviewers and supervisors requires an understanding of personal relationships likely to turn up during the course of an interview.

These ideas of objectives of sample design and of interviewing are only a part of the whole complex work of market research. They are, however, important division points along the right of way. In turn, each illustration which has been presented above raises its own peculiar technical questions as a problem in market research. For the purpose of getting an illustrative peek at the technical questions involved, it will be sufficient to select only the tape recorder problem and "The Jolson Story."

The objective of the tape recorder was to discover whether it possessed qualities which would make it unacceptable in the market. The technique of sample design was limited to the trial experience of a number of able secretaries who would try the instrument and by trial determine its acceptability. The interviewing part of the work was simple because the relatively few interviews were under the immediate supervision of a single individual in one city.

By contrast the research techniques for "The Jolson Story" were very much more complicated. There were in fact here a series of objectives so that the whole research represented a series of steps. Again assume that we are concerned only with the initial problem of the title choice. The objective then was to find the most acceptable title and to measure its acceptance in relation to the known acceptance of other titles of other moving pictures which had been on the market. This immediately implies comparison with standards obtained from other previous market studies. The sample design was that of a cross section of moving picture theater audiences classified by age, by geographical distribution, and by other characteristics. The sample used was representative of the people of the whole United States. Obviously it required a series of elaborate studies in order that it should be a truly representative cross section of this market. Finally the work of the data collection had to be accomplished through a group of interviewers. These carried on interviews in all cities with population of 50,000 and over as well as in other less densely populated centers. Supervision was obtained by mail correspondence and traveling supervisors. Thus every care was used to insure that the ballots from the public represented a true opinion.

Throughout this paper emphasis has been placed upon the importance of consumer attitudes in the setting of quality control targets. In the last analysis services as well as merchandise must be sold. The quality control engineer may assume that these problems are remoted from his immediate interest. This, however, is not the case. For unless he studies the market demand and understands it, the quality control engineer may be anxious to build in qualities which are not important.

In another but very significant sense, the quality control engineer also has much more limited markets much closer to his personal interests. These are the executives and the administrative officers of his company to whom he must sell his own product which is the service of quality control. If he is negligent in the attempt to learn what his executives have in mind, if he is negligent in the use of the evidence of what his service can perform, if he is negligent in the opportunities to see that the program he has planned is useful in advancing sales, his own work is at least haphazard if not fruitless. Setting quality control targets through market research is partly a problem of sales and of marketing, but it is equally true that it is a problem of the quality control engineer when he tries to sell his product in a somewhat more limited market represented by the executives of his own company.

APPLYING S.Q.C. IN THE BREWERY BOTTLE HOUSE
- TASTE UNIFORMITY & BOTTLING OPERATIONS -

Everett P. Hokanson

Blatz Brewing Company

Since May, 1951, Schenley Distillers (of which Blatz is a subsidiary) has been applying statistical principles in a Uniformity Taste Control Program and initiated a similar installation at Blatz in May, 1953. This initiated our S.Q.C. Office as a section within the Quality Control Department which includes the Laboratory for chemical, physical and bacteriological testing.

Our first concern in objective quality control of the finished product is uniformity of taste. Whatever has been predetermined as the most desirable brew from the consumer acceptance subjective view point, it is a primary concern of the Brewery to maintain taste uniformity for the bottled beer. This can not be accomplished by chemical tests and may be risky to leave to the subjective and variable judgement of a select few according to traditional methods.

However, it also follows that the physical factors such as air, gas and fill contents must be controlled to perpetuate the taste uniformity for tanks released to bottling. In this respect, our formal preparation of S.Q.C. charts for control of bottling operations started in May, 1954.

Part I Uniformity Taste Control Program

In so far as our samples for statistical control of taste uniformity are drawn from the Bottling Tanks, this control is supplementary to the taste uniformity previously checked by the Brewmaster's Taste Committee when the same beer was ready for transfer from the Finishing Cellars — however, still subject to final filtration and blending as well as further handling.

Our psychometric laboratory is centrally located in the Bottle House. It is equipped with a wall of 5 taste booths partitioned off from the room proper and entered by separate door to provide neutral taste test conditions. Samples are served the panelists from a service counter in the rear through an enclosed turn-table.

Originally we scheduled 20 panels of 5 employees each (100 tasters) for both morning and afternoon tests. A year later we selected the 50 best rated tasters and rescheduled for only 10 panels by altering our procedure to serve quadro-trio instead of duo-trio random sampling pattern. When this paper is delivered we may have again selected only 25 tasters to re-balance panels and sharpen the taste control.

(As of March '55) Our procedure is to draw 12 oz Production Samples from each bottling tank and prepare trays of 8 cc samples (2 oz glasses) for comparison with the Standard Sample - each test being a triad. Each tank is tested using the duo-trio pattern (2 tests of 3 glasses each) for each panelist; thus providing 10 tests per panel or 20 tests for 2 panels. With the quadro-trio pattern we may obtain test data on 2 tanks simultaneously. Usually only 20 tests (judgements) are necessary to pass each tank.

We apply statistical quality control to tasting as follows: (1)

Even though the Standard and Production Samples may be exceptionally close in taste, a person may correctly match a large number of samples by chance alone 50% of the time - unless he contradicts procedure. Therefore as in the "heads and tails" experiment, our Uniformity Mean is 50% as is best demonstrated by the frequency curve for panel scores obtained when selecting New Standard. (see Chart #1) In test situations where compared samples are less uniform, the mean point for average performance may move up the scale considerably, but this is an indication of discrimination in taste rather than chance results.

Panel scores (10 tests) may range from "0 right & 10 wrong" response (a 0-10 score) to "10 right & 0 wrong" response (a 10-0 score). However, we consider 3-7 scores or less and second 4-6 as invalid - which actually truncates the normal frequency curve at the lower end. However, having treated negative scores as invalid, we assure that significant high scores are not nullified by abnormally low scores. By serving another panel we can check which of the previous high or low scores most nearly reflected the true degree of uniformity. For Tank Acceptance, 2 panel scores range from 9-11 to 14-6. As a 15-5 score suggests possible significant difference, we serve a third panel in these instances and reject the tank if the 30 test score is correspondingly high. (this resembles a double sampling plan for attributes.) The rejected tank is split and each half topped with other beer in transfer. Both tanks are then subjected to retesting.

The <u>Standard Deviation</u> for frequency distribution of rating scores is determined as that of a normal binomial distribution as follows:

With $p' = 50\%$ (correct response)..(as to accuracy in matching samples
$q' = 50\%$ (incorrect response) (by detecting <u>taste difference</u>, thus
n = number of tests served (indicating degree of <u>non-uniformity</u>

And using the formula $\sigma p' = \sqrt{\frac{p'q'}{n}}$ compute Standard Deviation as:

For 2 panels: (20 tests) $\sqrt{\frac{.50 \times .50}{20}} = \underline{0.112}$ For 3 panels: (30 tests) $\sqrt{\frac{.50 \times .50}{30}} = \underline{0.0913}$

To construct a <u>scale for rating differences</u> between Standard and the Production Samples, we determine the number of σ by which each tank acceptance score differs from the position for $p' = 50\%$ under the normal curve. For example:

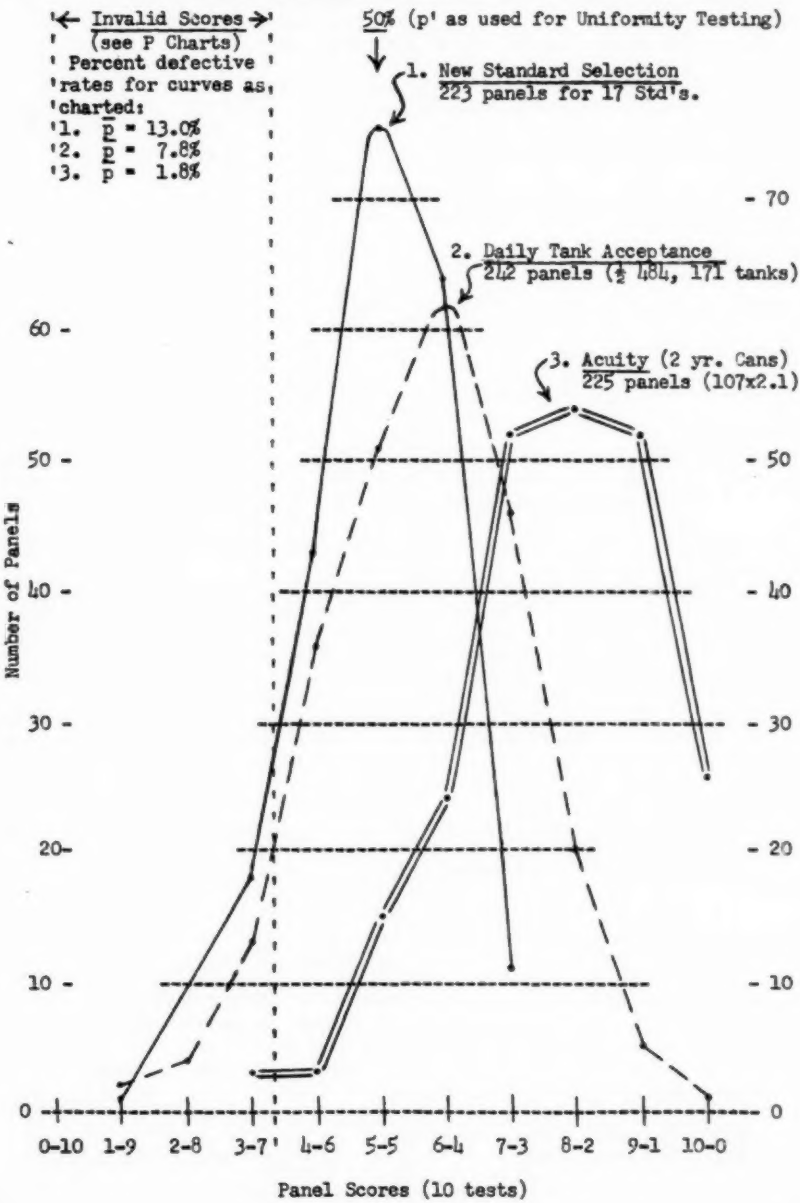
Using the formula: $x = \frac{\%\ \text{difference}}{\sigma p'} = \frac{p - p'}{\sigma p'}$ {which may be considered as $= t$ or z,

For 2 panels <u>scoring 15-5</u>: (equals 75% on 20 tests) $\frac{75\% - 50\%}{0.112} = +2.2\ \sigma$

For 3 panels <u>scoring 24-6</u>: (equals 80% on 30 tests) $\frac{80\% - 50\%}{0.0913} = +3.2\ \sigma$

Your choice of a 30 test score for rejection depends on the degree of assurance desired for uniformity and the confidence placed in the discriminatory ability of your panelists. You may elect to reject at a+3 σ or slightly lower position.

Chart #1 FREQUENCY DISTRIBUTIONS FOR VARIOUS TYPE TEST SITUATIONS



However, to determine the % of risk for needless rejection at a particular sigma position, you may wish to consider the bias introduced by eliminating invalid scores. This bias is more probable for close uniformity than known non-uniformity test situations. (Chart #1)

Entering the Tables for Cumulative Probabilities with $z = 2.15\ \sigma$, which is computed for a 21-9 score, we obtain 0.9842 (area under normal curve) or 98.4%. This leaves 1.6%, or the risk of 16 rejects out of 1000 that rejection was unnecessary.

However, to consider possible bias, we must use the Binomial Expansion term for probability expressed as:

$$C_r^n\, q'^{\,n-r} p'^{\,r}, \quad \text{or } C_r^n\,(0.5)^n$$

Solve for all possible combinations of r (# correct in 30 tests). Assuming p' as 50%, and developing data for three plans, we obtain % of risk as follows:

Assume decision point for r as 14.5; on 20 tests, accept if 14, retest 15
and 20.5; on 30 tests, accept if 20, reject 21

Plan 1. (no bias): always 30 tests, any score valid, ... % risk = 2.2%
2. (1 bias): accept on 20 & reject on 30 tests, ... % risk = 0.9%
3. (2 bias): same as #2 except 3-7 & 2nd 4-6 invalid; % risk = 1.6%

Therefore we can assume that 1.6% is the risk at 21-9 reject, if p' = 50%.

We obtain New Standard within every two weeks by matching samples of selected Bottling Tanks with Current Standard until achieving a 100 test score as close to 50-50 as obtainable, and not in excess of a 55-45 score, which is 55% correct. Using the formula stated previously, we find that the Standard Deviation for 100 tests is 0.05 and rates as 1 σ. As continuity of Standard is essential to the maintenance of taste uniformity, you can see that this sigma position is the maximum to be tolerated. This emphasizes the advantage of eliminating invalid scores so as to intensify the effect of 7-3 and higher panel scores in disqualifying doubtful tanks as New Standard.

Panel balance, consistent performance by individuals, strict control on maintenance of Standard, and minimum occurrence of invalid tests - are all vital factors in the success of this program.

A P chart for Invalid Tests is maintained daily for the purpose of observing factors contributing to defective scores. (see Chart #2) We compute p daily and post to a chart maintained directly over the serving counter. In this location the panel operators are constantly aware of their responsibility to minimize occurrence of invalid tests where controllable. Attitude, skill of tasters, timing, temperature and fill level variations, as well as contrasting test situations, may have considerable effect on the outcome of a panel score. The basic formulas for this fraction defective are as follows:

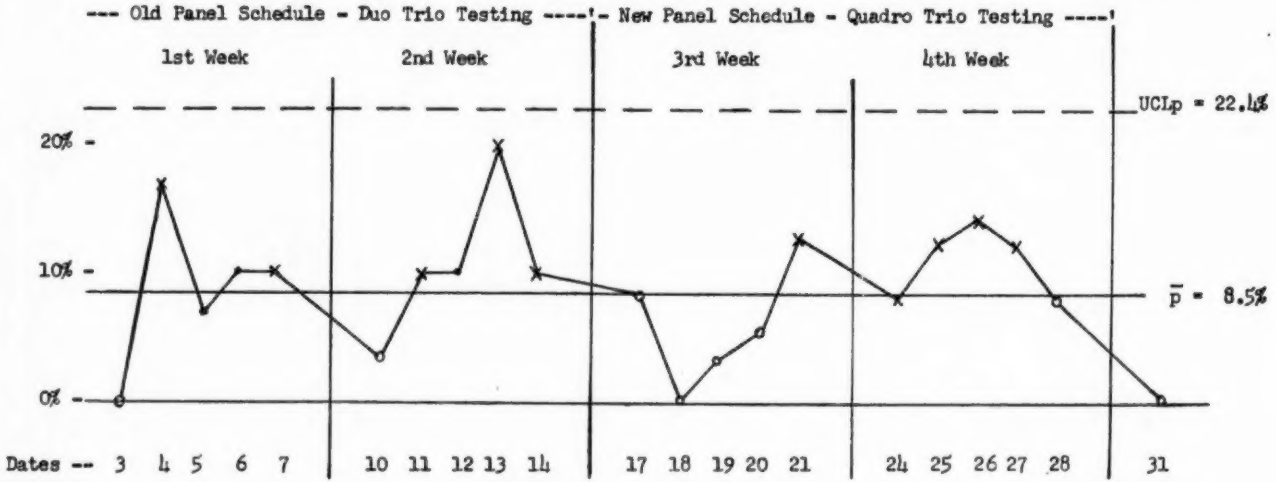
$\bar{p} = \dfrac{1\ (\text{\# of invalid scores, 3-7 or less})}{n\ (\text{\# of panels served})}$ $\qquad UCLp = \bar{p} \pm 3\sqrt{\dfrac{p\,(1-\bar{p})}{n}}$

(1) also see Schenley Q.C. Laboratory Procedure, issued by D. Brandt 7/52

Chart #2

P CHART for INVALID TESTS (Jan. '55)
(Panel scores of 3-7 or less)

x pts. - primarily Std. Search tests
o pts. - " Age Check "
. pts. - " Tank Acceptance



Analysis of Performance:

Test situations with respect to $\bar{p}$ for Jan.

Test Situation	Invalid Presumed Tendency	Points Above $\bar{p}$	Points On or Below $\bar{p}$	Number Panels (N)	Number Invalids (i)	Situation Percent Defective
Std.Search	Highest	9	1	150	25	16.6 %
Age Check	Lowest	0	8	276	10	3.9 %
Acceptance	Average	2	1	312	28	8.9 %
		11	10	738	63	8.5 %

Part II CONTROL CHART for BOTTLING OPERATIONS

As of this writing our $\bar{X}$ and R chart program is only 9 months old and is definitely still in the experimental stage of seeking the right design for practical control of a complicated multi-variable situation. There are many other SQC applications which I am not prepared to demonstrate at this time; however, I will welcome your comments during the discussion. Currently we are using a "three in one" $\bar{X}$ and R chart. The several inter-related as well as independent variables have required adding or deducting features periodically in the attempt to find the proper graphic relationship.

Although plotted daily, our charts cover a 4 week period and are thoroughly reviewed in open discussion by the Quality Control Committee within a few days after their issuance. This Committee includes the Exec. Vice President, Production Manager, Master Brewer, Bottling Superintendent and Assistant, Maintenance Superintendent, Chief Chemist and SQC Supervisor. Whereas, we previously issued a set of charts (9 lines) for each committeeman, we now contemplate issuing only one set for the V.P. and one file copy for projection on a screen during the discussion. To gain more immediate control, an hourly can and quart line air content report is posted in the Bottling Office with critical data for other lines noted as observed. This is supplemented the next morning by the Laboratory's full typed report for all tests and a weekly Data Sheet issued by SQC Lab summarizing all R and $\bar{X}$ data. Therefore, the 4 week charts enable us to observe the week to week trends for major policy decision, while hourly control is assured by prompt advisory service by both the Laboratory and S.Q.C. Office.

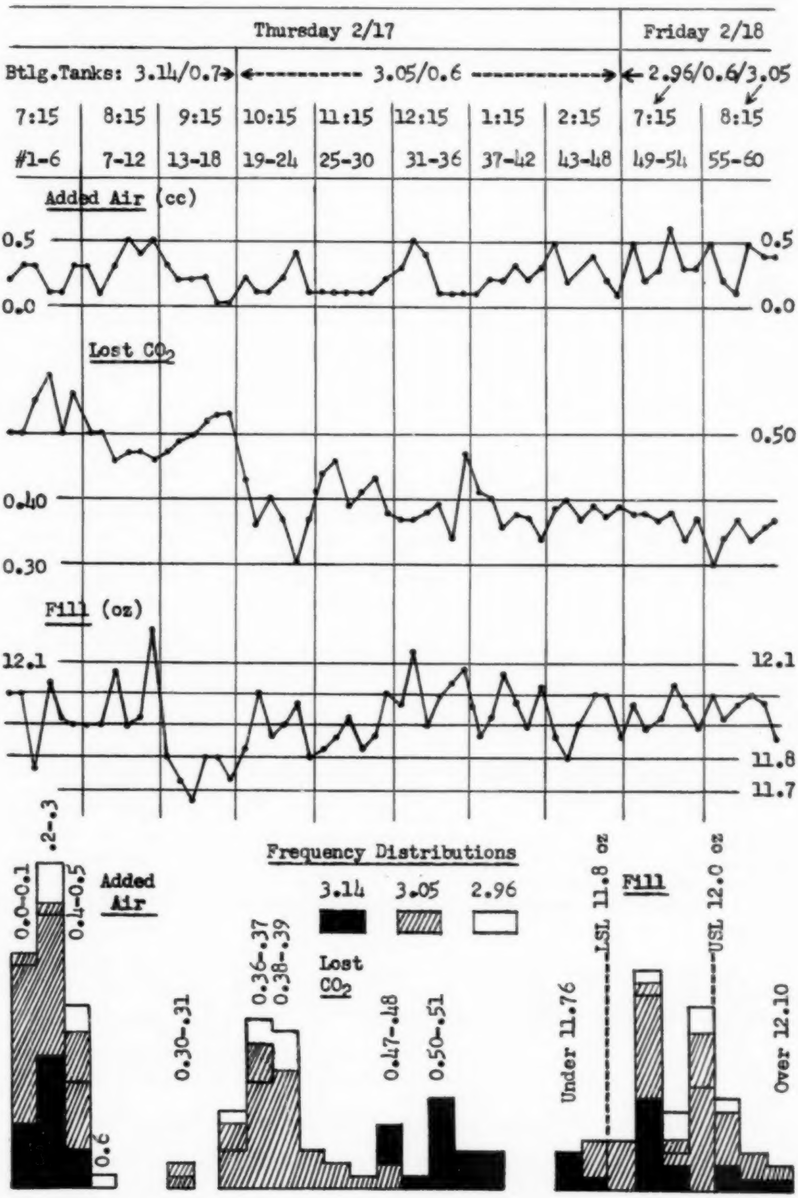
Several breweries, after analysis of this subject, have concluded on drastically modified control procedures such as controlling on the basis of frequent foam-over observations or restricting sampling to such critical periods as major filler stops. Your decision rests largely on the maximum average air content you desire, conclusions as to the head-to-head or within-head variation pattern for your fillers, and confidence in the control of air and CO_2 content during transfer from bottling tanks to the filler.

When considering $\bar{X}$ and R charts we first have certain basic facts to recognize in design of the sample plan so as to properly evaluate the test data. Testing for gas, air and fill is both destructive and time consuming, so we naturally desire the minimum practical sampling plan.

Whereas, an industrial lot acceptance plan evaluates samples as produced by a single machine, a filler is actually 50 to 60 different units (filler heads) operating in combination as one producing unit. Each of these units may produce defectives for an independent cause, temporarily or consistently. This will show up as abnormal Range on the chart but such a point is rarely repeated in a strictly random sampling plan with chance selection of samples off different filler heads. However, when several units produce similar abnormal results, the $\bar{X}$ chart reveals the situation and the cause may be traced to the beer received or malfunctioning of the filler as a whole.

With respect to head-to-head variation it is interesting to examine the frequency distributions and relation of Added Air, Lost Gas and Fills as revealed in Chart #3. Single samples were taken of 6 consecutive can filler heads each hour over a 10 hour period.

Chart #3 FILLER HEAD-TO-HEAD STUDY & Effects of Varying CO_2 Volumes 12 oz Can Line (60 spouts) Single Samples, Groups of 6



Although these were not simultaneous samples (except for subgroups of 6) and do not indicate a particular head's performance for successive tests, this chart does reveal the possible range of variation in the performance of 60 heads as well as the behavior of the filler to different cellar CO_2 conditions. It's evident that alertness for filler head adjustments is advisable for even a single abnormal R point on the chart rather than considering it a chance cause.

However, the relationship of these variables interests me particularly. If low fill contributed to high added air, why do many tests such as 11.76 to 11.85 oz. correspond to as little as 0.2 to 0.3 cc added air -- unless exceptional foam-over contributed to both results. But if a good foam-over minimizes the pick up of air we should usually find a corresponding larger amount of expended CO_2. A correlation proves this point for the low fills but often shows minimum loss of CO_2 for preferred fills even with the minimum pick up of air. To properly complete a correlation we need consideration of circumstances contributing to the "bound in" qualities of CO_2 and resultant foam-over; such as storage period in Finishing and Bottling Cellars while under final carbonation, the heavy or light character of the foam formation itself, and temperatures. This reveals the difficulty with judging control of air and fill based solely on foam-over observations, which requires expert judgement of all related results.

We realize that our plan of obtaining at random 3 samples hourly off each can and quart line filler and only 3 samples each 3 hours off each other bottle line provides an insignificant sample size by statistical standards as well as being considerably disproportionate for the varying high production rates of different lines. However, this is justified by the supplementary control aids provided. In consideration of minimum sampling coupled with head-to-head and within-head variation, we propose to alter our plan to require identification of filler heads sampled so as to trace abnormal range. By resampling a particular head within successive hourly samples (including two other heads) we can observe if the abnormal condition is corrected.

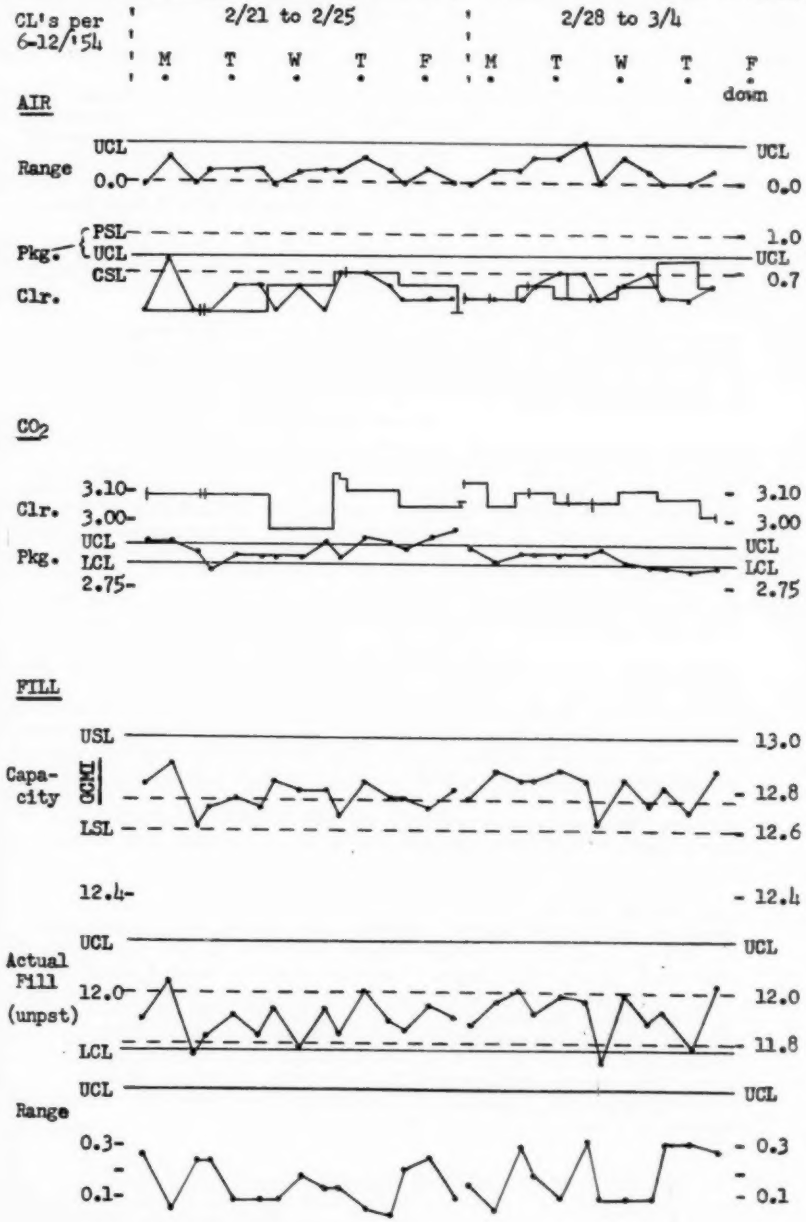
So the Control Chart may serve its purpose for confirmation of performance, trend comparison and policy decision (such as Spec. Limits), we rely heavily on the diligence of the filler operators and foremen to be on the alert for immediate corrective action by removing obvious defectives at the filler and making adjustments as indicated. Daily foam-over observations are made by a lab technician as a check on operator vigilance Theoretically, it should be rare that sampling will reveal critical conditions. When it does occur, it is frequently a situation not apparent to the operator and requires immediate inspection before assuming chance cause.

At this point we are prepared to examine a typical $\bar{X}$ and R chart as shown in Chart #4. First note the three principal divisions for Air, Gas and Fill. Each division is subdivided as follows:

Air: Range, Package & Cellar Air-(vertical spacing - Added or Reduced Air)
Gas: ..... Package & Cellar Gas-(vertical spacing - Expended CO_2)
(variable & level line)
Fill: Range, Actual Fill & Full Capacity (spacing - approx. Head Space)

Specification Limits are provided for Cellar and Package Airs (USL's), Full Fill Capacity (USL & LSL) per GCMI, and Fill Level (USL & LSL).

Chart #4 BOTTLING CONTROL CHART, Line #6, 12 oz bottle (3 sample average)



Control Limits (ruled in red) for Package Air, CO_2 and Fill are based on June-Dec., '54, 6 month $\bar{R}$ and computed by $A_2\bar{R}$ method.

(see Chart #4) 12 oz Bottle Line for 2 week period, Feb. 21 thru Mar. 4.

AIR: Range indicates that head-to-head variation was within control and averaged 0.1 & 0.1 cc for each week (per Data Sheets). Although on 2/28 & 3/1 there was a critical upward trend, correction is noted.

Cellar Air exceeded its Spec. Limit of 0.7 cc in only one instance. However, we note a tendency for Air to increase within each week.

Package Air averaged well below Spec. Limit of 1.0 cc as well as the UCL. Most tests indicated "same as" or reduced the Cellar Air. The weeks averaged 0.6 cc.

GAS: Cellar CO_2 for tanks tapped to this line averaged 3.07 and 3.09 for respective weeks. (Currently, Cellar CO_2 Spec. Limits are being set in relation to each line's use of CO_2 in filling so as to assure uniform Package Gas.)

Package CO_2 averaged 2.89 and 2.84 for each week, whereas, expended CO_2 averaged 0.18 and 0.25 volumes. This lower loss average the first week is seen to be credited entirely to Wednesday and Friday. The very narrow Control Limits (based on '54 $\bar{R}$) indicates tendency to drift, although on 3/3 the downward drift is traced to Cellar CO_2 as the assignable cause.

Fill: The Full Fill Capacity for bottles averaged quite closely to the GCMI spec. average of 12-23/32 oz. All bottles were Returnables.

Actual Fill averaged 11.90 & 11.91 oz. for the two weeks. The LCL of 11.75 is based on '54 $\bar{R}$ and lies outside the Spec. Limit. This is contrary to usual relationship of Control Limits but indicates a need for tightening the Range rather than altering Specs. However, the close positive correlation of Fill to Capacity suggests where a principal correction is needed — i.e., bottle standardization.

Whereas the samples are obtained off the filler, it may be presumed that an indication on the chart for short fill tendencies will have also been detected at final inspection.

Range of Actual Fill averaged 0.16 & 0.20 oz for each week and remained well below the UCL. However, we cannot treat this Range as largely a measure of head-to-head variation in the manner we evaluated Package Air Range. You will note a tendency for negative correlation with Actual Fill which we have already observed as correlating positively with Fill Capacity. This may be due to greater differences (Range) for measured liquid quantities "averaging" as low fills than the corresponding quantities for high fills due to differences in the bottle diameters at the respective points. However, a correlation analysis may be necessary to establish this statement.

CONTROL CHART ANALYSIS OF ENGINEERING EXPERIMENTS

Bonnie B. Small
Western Electric Company

In recent years, more and more of our engineers have been showing an interest in statistical design of experiment. At the same time, more and more of them have been learning how to make engineering capability studies with $\bar{X}$ and R charts. The kinds of information they get from their $\bar{X}$ and R charts are so valuable, for engineering purposes, that it is only natural that these engineers should be keenly interested in the use of $\bar{X}$ and R charts to analyze the results of their designed experiments.

At Western Electric we teach a course to our engineers on this subject, which is called "Control Chart Analysis of Engineering Experiments." It starts out with the basic principles of experimental design. After that the engineers are given a certain amount of practice in analyzing four and five factor experiments using the mean square method of analysis of variance. Then they learn how to analyze these same four and five factor experiments with control charts.

The following material is taken from this course. I have tried to include enough to give an idea of the speed and facility with which this technique can be used, without going into too many of the details.

In a five-factor experiment of the pure factorial type (each factor at two levels), the engineers are expected to be able to make all the necessary calculation in less than 15 minutes. It takes them from 10 to 20 minutes to plot a typical chart.

Control Chart Method

Step 1. Set up the data.

One of the examples used for practice work is shown in Figure I.

Step 2. Calculate $\bar{X}$ and R.

This is done very rapidly by filling in the form shown in Figures II and III. The instructions given to the engineers are as follows:

Take the data in pairs as directed on the form. In Figure II we are told to form horizontal pairs, and in Part 1 they are to be adjacent.

Starting at the upper left-hand corner of the data, we find that the first two values are 11 and 14. $\bar{X}$ is 12.5, and R is 3. Record these in the appropriate columns.

Then, going back to the data, move down one step and take the next horizontal adjacent pair. These are 5 and 8. $\bar{X}$ is 6.5, and R is again 3.

Continue taking the pairs downward as far as you can go, then move to the next section of the data and repeat the procedure until Part 1 is completed.

When you come to Part 2, continue taking horizontal pairs, but this time skip one. For example, starting again at the upper left-hand corner of the data, take 11 and 17. Then continue downward as before until all the pairs have been used.

In Part 3 take horizontal pairs again, but this time skip three. The first pair will be 11 and 8.

In Parts 4 and 5 follow a similar procedure, except that now you form the pairs vertically instead of horizontally.

Note that when Figures II and III are completed, you have calculated $\bar{X}$ and R values for all possible combinations of the variables.

<u>Note</u>: In doing this, the engineers are instructed to pay no attention to the identifying variables in the first five columns of the form. They merely fill in the data mechanically, and the $\bar{X}$ and R values fall in the proper places.

Step 3. <u>Obtain the residual (adjusted values of R).</u>

Since this is a designed experiment, we have deliberately introduced variables at different levels. Every difference in level is inflating our values of R. For this reason we do not use the R values directly, as we would in an ordinary R chart, but instead we use them indirectly as a means of removing the inflation.

To do this proceed as follows:

When you make your original calculations, record a plus or minus sign in front of each value of R. In the case of horizontal pairs, record a plus sign whenever the right-hand member of the pair is larger, and a minus sign whenever the right-hand member is smaller. In the case of vertical pairs, record a plus sign when the bottom member of the pair is larger, and vice versa.

Now take the algebraic average of the R values for any Part. This algebraic average will be equal to the <u>difference between the levels</u> of the variable summed across. For example, in Part 1, where we are summing across C, the algebraic average of the R values is +4. This means that C2 is 4 higher than C1.

Now look at the final column on the form, which is headed "R-crossed." The R̶ symbol stands for "adjusted value of R." These are the values of R we would have obtained if there had been no difference between level 1 and level 2.

To obtain these values rapidly, proceed as follows:

Take the algebraic average of the R column, which in the case of Part 1 is +4. Record in the R̶ column the difference between this algebraic average and each value of R.

In the first line of Part 1, the difference between +4 and

+3 is 1. In the third line of Part 1, the difference between +4 and -6 is 10.

It will be recognized that this is a rapid method of adjusting for systematic differences within subgroups, as outlined in References 1 and 2. When we use the algebraic average of the R's for one entire Part, we are adjusting the data for main effects only. But the same thing can be done in the case of interaction.

Step 4. Plot the charts.

We now make ordinary control charts for samples of $n = 2$, using the $\bar{X}$ values as usual and substituting the R_k values for R.

The samples can be plotted in any arrangement desired. For example, we might gather together all the samples representing A1 and compare them with the samples representing A2.

Step 5. Interpret the charts.

The $\bar{X}$ chart is read like any control chart for averages of samples of 2. The R_k chart is read like any control chart for ranges of samples of 2.

Since this is a designed experiment, however, it is necessary to keep in mind that we are looking at the same data formed into samples in many different ways. For example, in applying our tests for significance, we have to be careful that we apply them only to groups of independent samples.

Tests of Significance

The engineers use the same tests of significance that they have already learned to use in their engineering capability studies. For example, on the $\bar{X}$ chart, they mark a cross whenever they find

(a) a single point beyond $3\ \sigma$ from the centerline.
(b) 2 out of 3 independent points beyond $2\ \sigma$.
(c) 4 out of 5 independent points beyond $1\ \sigma$.
(d) 8 independent points in a row on one side of the centerline.

The tests are calculated in such a way that, if the parent population is normal, each test has roughly the same degree of significance.

In the case of the R_k chart, where the distribution of ranges for samples of 2 is not symmetrical, they use the same tests applied to slightly different areas. For example, the "2 out of 3" test applies to the upper one-half of the control band instead of the upper one-third.

Practical Helps

Very early in the use of this technique at Western Electric, it was discovered that the engineers would need a guide for rapid plotting. This was worked out. It is called "Combinations of Variables in a Five Factor Experiment." There is also a similar set of combinations for a four factor experiment. With the help of this guide the engineer is able to select, with great rapidity, the samples representing any desired

set of main effects or interactions.

The guide also shows him automatically which samples are independent.

As an illustration of how the guide is used, the following section covers the interactions between variables A and E. The engineer simply turns to the designated Part and plots in order the indicated samples.

A1E1

INTER.	PART	SAMPLES
B	4 1 4	1, 3 1, 2, 5, 6 5, 7
C	4 2 4	1, 5 1, 2, 5, 6 3, 7
D	1 2 1	1, 5 1, 5, 2, 6 2, 6

A1E2

INTER.	PART	SAMPLES
B	4 1 4	2, 4 3, 4, 7, 8 6, 8
C	4 2 4	2, 6 3, 4, 7, 8 4, 8
D	1 2 1	3, 7 3, 7, 4, 8 4, 8

A2E1

INTER.	PART	SAMPLES
B	4 1 4	9, 11 9, 10, 13, 14 13, 15
C	4 2 4	9, 13 9, 10, 13, 14 11, 15
D	1 2 1	9, 13 9, 13, 10, 14 10, 14

A2E2

INTER.	PART	SAMPLES
B	4 1 4	10, 12 11, 12, 15, 16 14, 16
C	4 2 4	10, 14 11, 12, 15, 16 12, 16
D	1 2 1	11, 15 11, 15, 12, 16 12, 16

Interpretation

Figure IV shows the data of Figure I, plotted in such a way as to bring out the interactions between variables A and E.

Figure V shows the same data plotted in such a way as to bring out the interactions between variables A and C.

The $\bar{X}$ chart of Figure IV is interpreted as follows:

> On the A1 chart, E2 is significantly higher than E1.
> The same thing is true on the A2 chart.
> Therefore, E2 is consistently higher than E1.
> There is an E main effect.

The $\bar{X}$ chart of Figure V is interpreted as follows:

> On the A2 chart, C2 is significantly higher than C1.
> This is not true on the A1 chart.
> Therefore, C2 is higher than C1, not consistently, but only in the presence of A2.

There is an A x C interaction.

In the same way, from the R chart, we get information about spread. The R chart in Figure IV is interpreted as follows:

On the A1 chart, E1 is more uniform than E2.
The same thing is true on the A2 chart.
Therefore, E1 is consistently more uniform than E2.
This is a main effect.

Furthermore, there is one measurement on the A1 chart which appears to be "wild", or quite different from the others. This measurement occurs under A1E2, in the B2, C1 and D2 sections. The wild reading is thus identified as A1B2C1D2E2.

In the data of Figure I, this measurement is "5".

The engineers enjoy going back to the original data and removing the known effects of changes in level, in order to see whether the control chart conclusions were correct. The original measurements from which we obtained the data of Figure I are shown in Figure VI. Note that the "wild" measurement, which was 5 in Figure I, is now -3.

They also enjoy analyzing the experiment by the mean square method, for comparison with the control charts. Almost invariably they find that they get more information from the charts.

The above discussion was based on a five-factor experiment, with each factor at two levels. The same approach can be used, however, for any number of factors and any number of levels.

Practical Example

The following is an actual experiment, designed and analyzed by a product engineer. This experiment was the work of Mr. Alex M. Hanfmann of the Western Electric Company. It was submitted, in slightly different form, as a term paper in one of his studies at Lehigh University. This was the engineer's first attempt at experimental design.

Manufacturing data relating to companies, materials, temperatures and solutions, and the names of the people who took part in the experiment, have been deleted or disguised.

Planning the experiment.

The problem involved a change in the appearance of certain glass parts after a series of chemical, heat-treating and assembly operations. The change in appearance was related to rather serious economic and quality problems. The trouble occurred in "batches" and would sometimes disappear completely for a while when one of the processing variables was changed. However, experiments of the usual type, involving the changing of one variable at a time, had produced only confusing results.

Many factors were suspected of contributing to this trouble. Questions of cost made it prohibitive to collect large amounts of data. The problem was further complicated by the fact that the trouble showed up as a visual indication only, and there was no means of measuring it objectively.

The engineer began by making a list of all the variables that were suspected of being able to influence this condition. He disposed of these one at a time by

(a) randomizing,
(b) holding constant, or
(c) including in the experiment.

See Figure VII for the way the engineer made these decisions.

The variables he finally selected were these:

1. Pre-annealing temperature.
   x degrees, x plus 50 degrees, x plus 110 degrees, x plus 125 degrees

2. Final annealing temperature.
   y degrees, and y plus 75 degrees

3. Cleaning solution.
   Old and new.

4. Assembly operation.
   Before and after assembly.

See Figure VIII for the way the experiment was designed.

Getting the units made.

The engineer issued special instructions and took special precautions in getting the units made, in order to protect the mathematical basis of the experiment. See Figure IX for typical instructions.

Measuring the unmeasurable.

He set up visual standards for rating the appearance of the glass.

See Figure X for method of visual rating.

He selected four engineers to serve as raters, and had them rate each one of the 16 specimens, before and after assembly. The raters did not always agree with each other, or with their own original ratings when they were asked to rate the same pieces a second time. Some of the raters insisted on recording to the nearest half.

See Figure XI for the ratings.

The engineer averaged the four ratings for each piece. Since he was interested only in the relative ratings, he multiplied the values by 8 to get more convenient numbers.

See Figure XII.

Making the analysis.

He now proceeded to analyze his results by plotting a series of control charts, using the methods described above. The first chart

showed that the best pre-annealing temperature was x plus 50 degrees.

See Figure XIII.

The second chart showed that a final annealing temperature of y degrees would result in trouble after assembly. Also, the new cleaning solution was far superior to the old.

See Figure XIV.

Conclusion.

This experiment shows that much can be accomplished by a control chart analysis of an experiment, even in cases where the effects we wish to study are very difficult to measure.

Advantages of Control Charts

In general, the advantages of the control chart method are those outlined in References 1 and 2.

1. The chart shows specifically which combinations of variables are high or low. In the case of more than two levels, it will also show trends.

2. The chart makes a direct comparison between averages at different levels.

3. The chart tests for control of variability as well as averages. It is possible to pick out a single subgroup or even a single measurement that is "wild" or out of control.

4. The control chart is easy to calculate and easy for ordinary people to understand.

5. It is possible to incorporate future data or the results of other experiments. Duplicate experiments can be compared to determine their consistency.

To this should be added that the control chart analysis is far more flexible than other methods. As one engineer expressed it, "You can do more with the data." Of equal importance is the additional stimulus to the engineering imagination, on which, in the final analysis, the solution of the problem depends.

These notes have stressed the mechanical details of the analysis, but the really important thing is "what you can do with the data." After seeing hundreds of engineering experiments, analyzed by various methods, I am convinced that you can do more with control charts.

References

1. "Quality Control Charts for $\bar{X}$ and R Adjusted for Within-Subgroup Pattern." A. E. R. Westman and B. H. Lloyd, Industrial Quality Control, March 1949.

2. "Stratification Control Charts." I. W. Burr and W. R. Weaver,

Industrial Quality Control, March 1949.

3. "Economic Control of Quality of Manufactured Product." W.A.Shewhart. D. Van Nostrand Company, Inc., New York. 1931.

FIVE FACTOR EXPERIMENT

		A1				A2			
		B1		B2		B1		B2	
		C1	C2	C1	C2	C1	C2	C1	C2
E1	D1	11	14	17	15	8	19	12	17
	D2	5	8	11	9	2	13	6	11
E2	D1	25	19	34	26	22	35	17	31
	D2	29	18	5	28	10	20	28	23

FIGURE I

HORIZONTAL PAIRS

PART 1. ADJACENT.

VARIABLES					$\bar{X}$	R	R_a	
A1	B1	-	D1	E1	12.5	+ 3	1	1
A1	B1	-	D2	E1	6.5	+ 3	1	2
A1	B1	-	D1	E2	22.0	- 6	10	3
A1	B1	-	D2	E2	23.5	-11	15	4
A1	B2	-	D1	E1	16.0	- 2	6	5
A1	B2	-	D2	E1	10.0	- 2	6	6
A1	B2	-	D1	E2	30.0	- 8	12	7
A1	B2	-	D2	E2	16.5	+23	19	8
A2	B1	-	D1	E1	13.5	+11	7	9
A2	B1	-	D2	E1	7.5	+11	7	10
A2	B1	-	D1	E2	28.5	+13	9	11
A2	B1	-	D2	E2	15.0	+10	6	12
A2	B2	-	D1	E1	14.5	+ 5	1	13
A2	B2	-	D2	E1	8.5	+ 5	1	14
A2	B2	-	D1	E2	24.0	+14	10	15
A2	B2	-	D2	E2	25.5	- 5	9	16
						+ 4.0		

PART 2. SKIP ONE.

VARIABLES					$\bar{X}$	R	R_a	
A1	-	C1	D1	E1	14.0	+ 6	4	1
A1	-	C1	D2	E1	8.0	+ 6	4	2
A1	-	C1	D1	E2	29.5	+ 9	7	3
A1	-	C1	D2	E2	17.0	-24	26	4
A1	-	C2	D1	E1	14.5	+ 1	1	5
A1	-	C2	D2	E1	8.5	+ 1	1	6
A1	-	C2	D1	E2	22.5	+ 7	5	7
A1	-	C2	D2	E2	23.0	+10	8	8
A2	-	C1	D1	E1	10.0	+ 4	2	9
A2	-	C1	D2	E1	4.0	+ 4	2	10
A2	-	C1	D1	E2	19.5	- 5	7	11
A2	-	C1	D2	E2	19.0	+18	16	12
A2	-	C2	D1	E1	18.0	- 2	4	13
A2	-	C2	D2	E1	12.0	- 2	4	14
A2	-	C2	D2	E2	33.0	- 4	6	15
A2	-	C2	D2	E2	21.5	+ 3	1	16
						+2.0		

PART 3. SKIP THREE.

VARIABLES					$\bar{X}$	R	R_a	
-	B1	C1	D1	E1	9.5	- 3	3	1
-	B1	C1	D2	E1	3.5	- 3	3	2
-	B1	C1	D1	E2	23.5	- 3	3	3
-	B1	C1	D2	E2	19.5	-19	19	4
-	B1	C2	D1	E1	16.5	+ 5	5	5
-	B1	C2	D2	E1	10.5	+ 5	5	6
-	B1	C2	D1	E2	27.0	+16	16	7
-	B1	C2	D2	E2	19.0	+ 2	2	8
-	B2	C1	D1	E1	14.5	- 5	5	9
-	B2	C1	D2	E1	8.5	- 5	5	10
-	B2	C1	D1	E2	25.5	-17	17	11
-	B2	C1	D2	E2	16.5	+23	23	12
-	B2	C2	D1	E1	16.0	+ 2	2	13
-	B2	C2	D2	E1	10.0	+ 2	2	14
-	B2	C2	D1	E2	28.5	+ 5	5	15
-	B2	C2	D2	E2	25.5	- 5	5	16

FIGURE II

VERTICAL PAIRS

PART 4. ADJACENT.

VARIABLES					$\bar{X}$	R	R_x	
A1	B1	C1	-	E1	8.0	- 6	0	1
A1	B1	C1	-	E2	27.0	+ 4	10	2
A1	B1	C2	-	E1	11.0	- 6	0	3
A1	B1	C2	-	E2	18.5	- 1	5	4
A1	B2	C1	-	E1	14.0	- 6	0	5
A1	B2	C1	-	E2	19.5	-29	23	6
A1	B2	C2	-	E1	12.0	- 6	0	7
A1	B2	C2	-	E2	27.0	+ 2	8	8
A2	B1	C1	-	E1	5.0	- 6	0	9
A2	B1	C1	-	E2	16.0	-12	6	10
A2	B1	C2	-	E1	16.0	- 6	0	11
A2	B1	C2	-	E2	27.5	-15	9	12
A2	B2	C1	-	E1	9.0	- 6	0	13
A2	B2	C1	-	E2	22.5	+11	17	14
A2	B2	C2	-	E1	14.0	- 6	0	15
A2	B2	C2	-	E2	27.0	- 8	2	16
						-6.0		

PART 5. SKIP ONE.

VARIABLES					$\bar{X}$	R	R_x	
A1	B1	C1	D1	-	18.0	+14	2	1
A1	B1	C1	D2	-	17.0	+24	12	2
A1	B1	C2	D1	-	16.5	+ 5	7	3
A1	B1	C2	D2	-	13.0	+10	2	4
A1	B2	C1	D1	-	25.5	+17	5	5
A1	B2	C1	D2	-	8.0	- 6	18	6
A1	B2	C2	D1	-	20.5	+11	1	7
A1	B2	C2	D2	-	18.5	+19	7	8
A2	B1	C1	D1	-	15.0	+14	2	9
A2	B1	C1	D2	-	6.0	+ 8	4	10
A2	B1	C2	D1	-	27.0	+16	4	11
A2	B1	C2	D2	-	16.5	+ 7	5	12
A2	B2	C1	D1	-	14.5	+ 5	7	13
A2	B2	C1	D2	-	17.0	+22	10	14
A2	B2	C2	D1	-	24.0	+14	2	15
A2	B2	C2	D2	-	17.0	+12	0	16
						+12.0		

FIGURE III

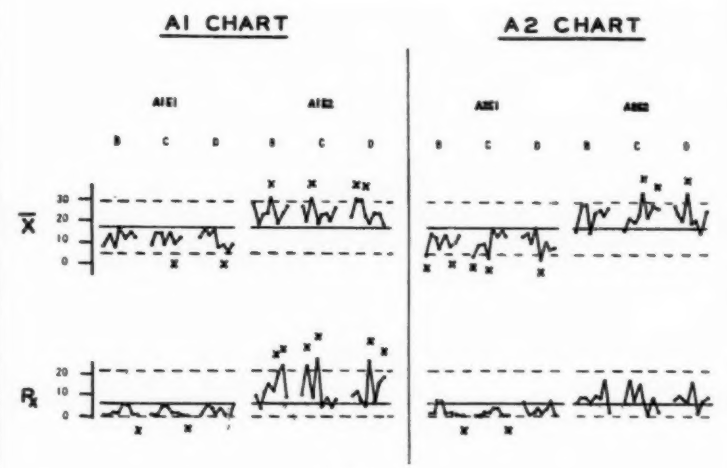


FIGURE IV

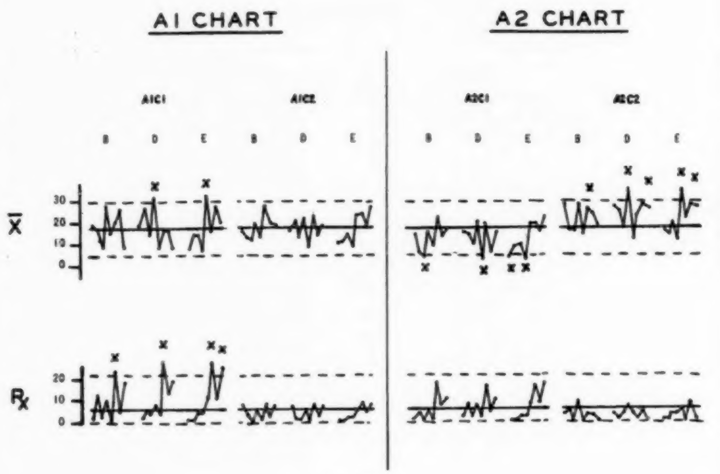


FIGURE V

ORIGINAL DATA

		A1				A2			
		B1		B2		B1		B2	
		C1	C2	C1	C2	C1	C2	C1	C2
E1	D1	11	14	15	13	12	15	14	11
	D2	11	14	15	13	12	15	14	11
E2	D1	13	7	20	12	14	19	7	13
	D2	23	12	−3	20	8	10	24	11

TO OBTAIN THE DATA OF FIGURE I, SUBTRACT 4 FROM THE A2C1 VALUES AND ADD 4 TO THE A2C2. THEN ADD 2 TO THE B2 VALUES AND 12 TO THE E2. FINALLY SUBTRACT 6 FROM D2.

FIGURE VI

Choice of Variables and Design of Experiment

Although a great many variables could have entered into the picture, it was decided to choose a four-factor experiment. This meant selecting the variables that appeared to be most important. This selection was done as follows:

Glass itself might be different from piece to piece. The only manufacturer of this glass pours the tubing only 2 or 3 times a year. It would not be feasible to change this material on short notice. This variable was excluded, and an attempt was made to randomize.

Chemical cleaning solution was the chief variable to be investigated (old vs. new.) This was included.

Temperature of solution was supposedly constant and not a variable, if the manufacturing instructions were followed. Excluded.

Composition of "old" cleaning solution. This solution was replaced in the tank only about once a week and was used for cleaning different metals. Conceivably, traces of different metals could react differently with acid and glass. It was decided to "randomize" this variable by using the "old" solution at the end of the week, so that all sorts of metal traces would be represented more or less randomly.

Heating after cleaning. In an engineering sense, this was not a variable, since heating had to be done in order to assemble the product. Statistically, the presence or absence of heating was one of the important variables, since it was known to affect the appearance. Included.

Annealing and pre-annealing. That the previous heat treating history of the glass could be responsible for what was happening after cleaning was considered ridiculous by some of the shop personnel. The writer, however, was looking for anything that could solve the trouble without a costly change in the glass type. Pre-annealing of the glass as received would have been quite cheap, and shifting the normal annealing of the glass to a different temperature would have meant no cost at all. Also, some textbook data indicated that investigation was justified. So both pre-

FIGURE VII

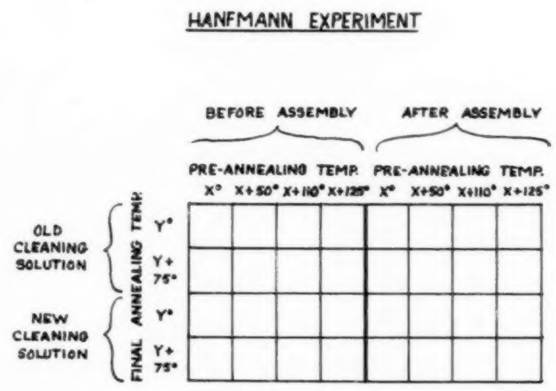


FIGURE VIII

INSTRUCTIONS FOR MAKING AN EXPERIMENTAL LOT OF GLASS PARTS

Mr. Z and Mr. W have a lot of pre-treated glass tubing. This lot is subdivided into 4 groups and comprises 27 pieces.

Group "P1" is marked with one file mark.	(6 pieces)
Group "P2" is marked with two file marks.	(6 pieces)
Group "P3" is marked with three file marks.	(6 pieces)
Group "P4" is marked with four file marks.	(9 pieces)

The identity of <u>each single piece</u> must be kept through all the subsequent steps. Therefore all operations must be observed by Mr. Z or Mr. W and at the end of each step every piece must be identified by the engineer, using file marks, etching or reliably affixed tags.

<u>Step 1</u>. Wash and dry glass tubing. Use clean water.

<u>Step 2</u>. Make convolutions as per layout. Take care not to break glass and not to lose identification by file marks.

<u>Step 3</u>. Anneal as follows:
One half of the pieces <u>in each group</u> to be annealed at y°C; the other half at y+75°C.
Annealing at y°C is designated "R1".
Annealing at y+75°C is designated "R2".

After annealing <u>each piece</u> to be marked according to its file marks and annealing temperature:

P1-R1 or P1-R2	P3-R1 or P3-R2
P2-R1 or P2-R2	P4-R1 or P4-R2

<u>Step 4</u>. Trim formed tubing to length. Make quite sure that the above marks (P1-R1 etc.) are transferred to trimmed length. Best by etching.

<u>Step 5</u>. Brush and wash insulators with soap solution, rinse and dry in hot air. Make seals for all insulators. Transfer markings to metal seals, using

FIGURE IX

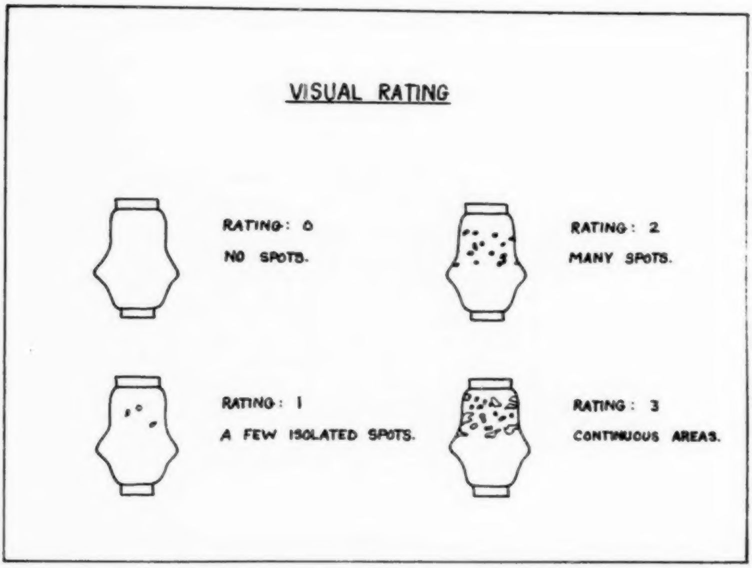


FIGURE X

HANFMANN EXPERIMENT

RATINGS AFTER ASSEMBLY

PRE-ANNEALING TEMPERATURE

	FINAL ANNEALING TEMP.	X° A	X° B	X° C	X° D	X+50° A	X+50° B	X+50° C	X+50° D	X+110° A	X+110° B	X+110° C	X+110° D	X+125° A	X+125° B	X+125° C	X+125° D
OLD CLEANING SOLUTION	Y°	2	2	X	X	1	3	2	5	$1\frac{1}{2}$	3	2	5	1	$2\frac{1}{2}$	1	2
	Y+ 75°	2	$2\frac{1}{2}$	2	2	$1\frac{1}{2}$	$\frac{1}{2}$	1	1	1	$1\frac{1}{2}$	0	2	$1\frac{1}{2}$	$\frac{1}{2}$	0	1
NEW CLEANING SOLUTION	Y°	0	$2\frac{1}{4}$	2	2	1	1	1	1	1	$2\frac{1}{2}$	2	1	1	3	$1\frac{1}{2}$	2
	Y+ 75°	1	$\frac{1}{2}$	1	1	$1\frac{1}{2}$	0	1	0	$\frac{1}{4}$	$1\frac{1}{2}$	1	1	$\frac{1}{2}$	1	0	0

FIGURE XI

HANFMANN EXPERIMENT

		BEFORE ASSEMBLY				AFTER ASSEMBLY			
		PRE-ANNEALING TEMP.				PRE-ANNEALING TEMP.			
	FINAL ANNEALING TEMP.	X°	X+50°	X+110°	X+125°	X°	X+50°	X+110°	X+125°
OLD CLEANING SOLUTION	Y°	8	6	12	10	16	18	19	13
	Y+ 75°	14	7	10	7	18	8	10	6
NEW CLEANING SOLUTION	Y°	10	6	8	15	12	8	13	15
	Y+ 75°	8	8	5	14	7	5	11	3

FIGURE XII

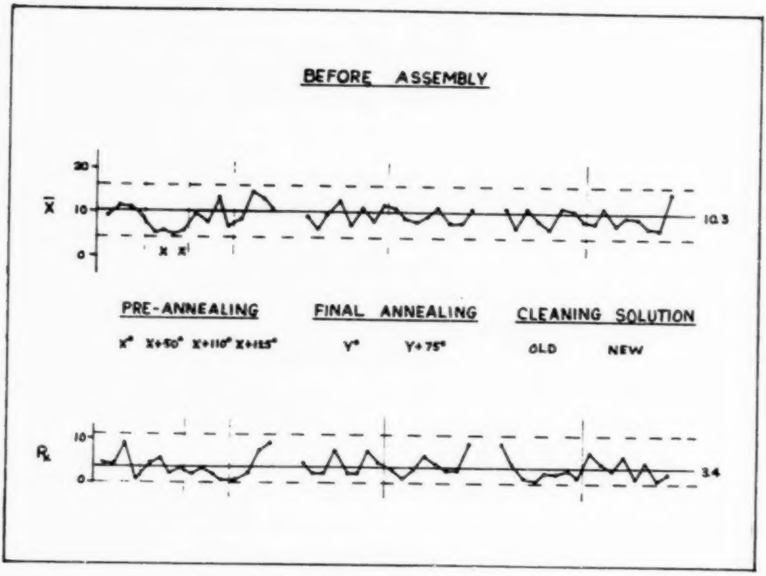


FIGURE XIII

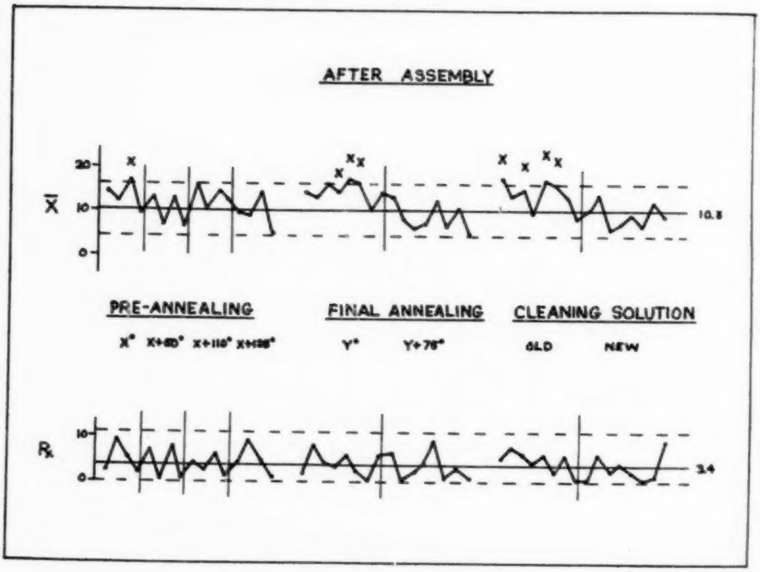


FIGURE XIV

ATTRIBUTES CHARTS: INTRODUCTION AND DEMONSTRATIONS

Max Astrachan
USAF Institute of Technology
and
Andrew S. Schultz, Jr.
Cornell University

Frequently in industry it is desirable to inspect and reject or accept product by classifying each item as to whether it is satisfactory or unsatisfactory with regard to the quality characteristic involved. This is termed inspection by "attributes." Inspection by means of limit or go and no-go gages is an example. Other examples are provided by inspections which judge a unit off-color, or scratched, or not full. Attributes inspection takes place for a number of reasons, generally because it is cheaper than actually measuring with micrometers or other measuring instruments, or because it is not possible to measure, as in the determination that an item is cracked or not cracked.

In such cases $\overline{X}$ and R charts cannot be used and a different type of control chart is available. This chart, while easier to construct and understand, is similarly founded upon the principles of mathematical statistics and works for the same reasons and in a similar fashion. Some special terminology is involved. The charts are termed "attributes charts" since the product is inspected by attributes.

There are three kinds of attributes charts in common use and it is well to be familiar with each. The first to be discussed is the "fraction defective" or p chart. Here the proportion or fraction of the product not conforming to the specifications is designated as p and used as a basis for the chart. This is perhaps the most frequently used attributes chart. A second type of chart is similar in conception but plots the "number defective" rather than the "fraction defective." The number defective is designated as "np" and is the number of rejected items found in a given number inspected. The third type of chart plots "defects per unit" which term is designated by the symbol "c." This type of chart is used in cases where there either is opportunity for a great many defects in each unit, or where the unit inspected is defined quite arbitrarily. An example of the first might be the number of defects in an assembly or flaws in a piece of glass, while an example of the second might be the inspection of cloth, paper, or wire, where the unit chosen may be a roll or an hour's production, or a piece of a given size selected in a prescribed manner.

The attributes control charts are used very effectively both in situations where 100% inspection takes place as well as in those where sampling is used. In fact, many firms took their first steps in statistical quality control by the introduction of attributes charts to situations where 100% inspection provided the data. In order to do this it is necessary to record not merely the number of defects or percent defective as almost all inspection processes do, with or without statistical quality control, but also to record the number or volume actually inspected. For the results of the chart to be useful in the control of quality it is necessary that information also be available concerning the material inspected. The type of information required is of course dependent upon the specific situation and purpose of the chart, but in general it is advisable to relate each quantity inspected to the source of production,

the time of production, and perhaps certain other factors such as material lot, process characteristics, or items which may be of importance in identifying possible causes of undue variations or poor quality.

Where sampling inspection is utilized a number of bases for selection of the sample size are possible. Frequently an acceptance sampling plan is in use and the data from the plan is used in judging process control. In other situations the sample size is determined by economic conditions and past experience. Always the sample size will be larger than that required for a variables or $\bar{X}$ and R chart, and in all cases it should be taken in such a fashion as to be representative of the process or procedure being judged.

As for $\bar{X}$ and R charts each of the attributes charts has a center line and control limits. To construct a fraction defective or p chart, we begin by collecting some data on the number of items inspected and the number found defective in each lot or sample or subgroup. The variable to be plotted is the ratio p, where

$$p = \frac{\text{number of defective items found in the sample inspected}}{\text{total number of items in the sample inspected}}$$

The center line is placed at the average value $\bar{p}$, where

$$\bar{p} = \frac{\text{total number of defective items found in all samples}}{\text{total number of items inspected in all samples}}$$

The upper and lower control limits are found from the following formulas:

$$UCL_p = \bar{p} + 3\sqrt{\frac{\bar{p}\,(1 - \bar{p})}{n}}$$

$$LCL_p = \bar{p} - 3\sqrt{\frac{\bar{p}\,(1 - \bar{p})}{n}}$$

where n is the number of items in each sample. We assume this is constant for all subgroups for the time being.

In order to see how this works we shall simulate a controlled production process by a box of beads, and take samples from it in a random manner. The box will contain a certain number of red beads which represent defective items, and white beads representing non-defective items. The results of the sampling will be analyzed and plotted on an ordinary p chart, and then interpreted as though the data represented an actual process. Sampling experiments of this type have been found very useful to illustrate the principles of statistical quality control.

For our first demonstration the bead box will contain a high percentage of red beads (defective items). Twenty samples of 50 each will be drawn and the results analyzed. This will involve the calculation of $\bar{p}$, UCL_p and LCL_p, using the formulas given above. These together with the fraction defective (p) for each sample will be plotted on a control chart. An analysis of the chart will show that because of the high percentage of defectives, under actual working conditions some type of major change will probably have to be made.

Assuming then that corrective action has been taken, our second bead box will be used to represent the new situation. The entire procedure used with the first box will now be repeated and the results analyzed. It will be noted that the average fraction defective is now considerably lower. Let us suppose that under actual working conditions the cost of further changes is prohibitive. If then, agreement is reached to accept this quality material, the resulting average fraction defective is adopted as a standard and called p'. The center line and control limits of the second demonstration can then be extended for future plottings. As for $\overline{X}$ and R charts then, as long as plotted points stay inside the control limits we assume a constant cause system is operating and conclude our standard product fraction defective (p') is being met by the process. If not, we have reason to suspect the presence of assignable causes and should take necessary steps to investigate the process and make whatever changes we can.

Although we used a fixed sample size in the demonstrations, and this is preferred in practical situations, occasions may arise in which this is not feasible. Thus lots may be reaching us which are of variable size and we may be inspecting them 100%; or the inspection may take place for all pieces produced by a shift or in one day. In such cases n, the number inspected, will vary from one subgroup to another, and a quite satisfactory procedure is to replace n by $\overline{n}$, the average value of n, in our formulas. In general the limits calculated this way will be valid as long as the correct sample size does not differ from the average by more than about 30%.

When the sample size is constant, it is possible to construct a number defective or np chart. Here we plot the number of defectives found in each sample instead of the fraction defective. For analyzing past data the center line is placed at $n\overline{p}$, the average number of defectives found in a series of samples and is given by

$$n\overline{p} = \frac{\text{total number of defectives found in all samples}}{\text{total number of samples inspected}} .$$

The formulas for the control limits are then

$$UCL_{np} = n\overline{p} + 3\sqrt{n\overline{p}\,(1 - \overline{p})}$$

$$LCL_{np} = n\overline{p} - 3\sqrt{n\overline{p}\,(1 - \overline{p})}$$

If a standard number of defectives np' has been found acceptable, we put the center line at this value and find the control limits by replacing $\overline{p}$ by p' in the above.

The data used in the bead box demonstration for the construction of p charts can also be used to illustrate the construction of np charts.

The third type of attributes chart, the c chart, is useful in situations in which classification of a product as defective or non-defective is an insufficient measure of quality. Thus an assembly could contain a large number of defects, and if we merely classified it as being defective we could not tell if it had 1 defect or 30 defects. We would not be using all the information available to us to improve the quality of our product.

Instead of a p chart, it is appropriate to use a c chart in such cases, where c is the number of defects per unit. The unit may be for example, an assembly, an area as in the case of a sheet of glass or cloth or paper, a length as in the case of a roll of wire, etc. The important thing is that the unit or "area of opportunity" for the occurrence of a defect must be defined and kept constant. Once this has been done we count for each such unit the number of defects, c, and plot these values on a chart. The center line is placed at $\bar{c}$, the average number of defects per unit, i.e.,

$$\bar{c} = \frac{\text{total number of defects found in all units inspected}}{\text{total number of units inspected}}$$

The upper and lower control limits are found from the following formulas:

$$UCL_c = \bar{c} + 3\sqrt{\bar{c}}$$

$$LCL_c = \bar{c} - 3\sqrt{\bar{c}}$$

The analysis and interpretation of c charts is similar to that of the other two attributes charts.

The attributes charts once initiated may be continued for a variety of purposes. If a process is being charted, and it is not in control, the chart can be continued until it is brought into control. In general this will be at a lower fraction defective than initially was the case. Periodically, the chart should be revised to recognize any change in quality level or improved conditions. Once the process is in control, the continuance of the chart can accomplish a number of purposes including (1) continued surveillance of the process to point up any out-of-control points and improved performance, (2) a continuing check on inspection consistency as evidenced by unusually high or low points that cannot be verified on reinspection, and (3) in the event of 100% inspection to provide information on consistency for the introduction of a sampling inspection procedure.

If sampling inspection procedures furnish the basis for the control chart data, the data may be used to develop evidence of process control at levels which indicate better than minimum performance and thus furnish a basis for reduced sampling and less inspection. Similarly, a continuation of a chart which demonstrates lack of control may be used as evidence for more rigorous sampling or 100% inspection procedures. Classification of lots of material by source and time becomes very important in the achievement of such objectives.

In conclusion, it may be stated that the use of attributes control charts can in general at a minimum cost provide great insight into the rudiments of statistical quality control and provide a basis to indicate where more effort in the application of variables charts or standard sampling plans may be economical.

REFERENCES

Although material on attributes charts is available in many places we list three useful books.

1. Burr, I. W., Engineering Statistics and Quality Control, 1953, McGraw-Hill Book Company, Inc., N.Y.C.

2. Grant, E. L. Statistical Quality Control, 2nd edition, 1952, McGraw-Hill Book Company, Inc., N.Y.C.

3. Duncan, A. J., Quality Control and Industrial Statistics, 1952, Richard D. Irwin, Inc., Chicago, Ill.

RELIABILITY OF GUIDED MISSILES

Robert Lusser
Redstone Arsenal

INTRODUCTION

Once when President Coolidge returned to the White House from a church service, Mrs. Coolidge asked him what the sermon was about. He answered: "Sin." She then asked, "What did the minister have to say about sin?" Coolidge answered, "He was against it."

This story illustrates one of the psychological aspects of the reliability problem. If you ask a thousand designers and manufacturers of missiles and missile components how they feel about the sin of unreliability, you will find that everyone is emphatically against it, yet you will scarcely find two people with the same opinion on what constitutes reliability and how it can be achieved. In particular, you will find that entirely too much faith is placed in time-honored standards of quality and reliability, which are obsolete so far as guided missiles are concerned.

1. RELIABILITY OF GUIDED MISSILES - A UNIQUE PROBLEM

Let me first discuss the general assumption that guided missiles are simply aircraft without a pilot -- a misconception that often leads to the erroneous conclusion that the components of guided missiles need not be as reliable as those of piloted aircraft, since no human being is aboard and at stake.

The basic error lies in a failure to consider the fact that in piloted aircraft only a dozen or so components, such as a wing, a stabilizer, and similar structural parts, are absolutely vital in that their failure would inevitably cause a loss of the aircraft. These few components are, of course, designed to be extremely reliable. There are, however, thousands of other components that are not vital because the pilot can parallel them in the event of failure, or do without them entirely until the aircraft is brought home for inspection and repair. These nonvital components, particularly all electronic components, thus need not be, and usually are not, very reliable.

In guided missiles things are basically different. There are not just a few components that must be considered vital. All components, down to the last relay, valve, or even soldered joint, are vital because the failure of any one of them will, with absolute certainty, cause the missile to miss the target. Since missiles that miss the target cannot be recovered and reused, they must be considered a complete loss. Loss of the missile alone might be valued at $100,000 and more, quite aside from the possible military disaster and loss of life which could result from its failure.

Let us imagine for a moment that in piloted aircraft the failure of any component, any tube, any relay, any soldered joint, would cause a catastrophe. Probably there would be no aviation at all, because no one would dare fly in such a deathtrap.

Everyone who is brought to appreciate this profound difference between piloted aircraft and guided missiles will at once realize that the

components of guided missiles must be made much more reliable than those used in piloted aircraft. The question arises: How much more reliable? To discuss this question, the mathematical aspect of reliability must be considered.

2. DEFINITION OF "RELIABILITY," "OVERALL RELIABILITY," AND "COMPONENT"

Before we can take a look at the mathematical aspect of reliability we have first to make some definitions.

a. Reliability: Reliability is sometimes defined as ability of a device to perform as prescribed. This, however, is incorrect. Reliability is not an "ability," but a probability, namely the probability that a device will perform as prescribed under all service conditions. Reliability must therefore be clearly understood as a mathematical term.

The proper definition of reliability is as follows:

"Reliability of a device is the probability, p, that it will function successfully under all environmental conditions occurring in service."

It should be noted that the time factor is omitted in this definition. Actually, the period of time during which a component is intended to operate perfectly is just one of the many dozens of design criteria and "environmental conditions," which must be considered in design, specifications, and tests. In guided missiles these other design criteria, for example, maximum shocks and temperatures might constitute far greater hazards to reliability than time of operation which is extremely short as compared to the required service life of normal equipment such as airplanes or radar.

b. Overall Reliability: If we evaluate the results of a number of missile firings we will find that a certain percentage will have been able to hit the target. This percentage is called the "overall reliability" of the missile type.

This overall reliability is one of the most important yardsticks of the military value of a guided missile type. We should, therefore, rely on it only if it is based on a statistically significant number of firings, say 20 or 30, at the least. To compute the overall reliability on the results of five or ten firings, will most certainly lead to illusions and wrong decisions.

c. Component: There exists an important mathematical relationship between the overall reliability of a missile type and the reliability of its components. Before we can discuss this relationship we must first define the term "component." A great deal of confusion prevails here. Some like to define as components whole packaged units which in themselves are highly complex; whereas others define even small and simple parts as components.

In order to lay the foundation for a sound philosophy of reliability the following definition of the term component is offered:

> "A component is an item that can be removed from an assembly. It is, however, not normally subject to further disassembly." Examples: A vacuum tube, a relay, a gyro, a wing, a servo motor.

Now it is obvious that the overall reliability of a missile type must somehow depend on the reliability of the components of which it is composed. But how?

3. MATHEMATICAL ASPECTS OF RELIABILITY

Simple mathematics of probability states that the overall reliability equals not the average, as some may believe, but rather the product of the reliabilities of the individual components:

$$P_{overall} = p_1 \cdot p_2 \cdot p_3 \cdots p_n$$

where p_1, p_2, p_3, etc., are the individual reliabilities of each of the n components.

This simple reliability formula is based on the following basic rule of probability:

> "If p_1 is the probability that an event, E, will occur and p_2 is the probability that an event E_2 will occur, then $p_1 \cdot p_2$ is the probability that both events will occur."

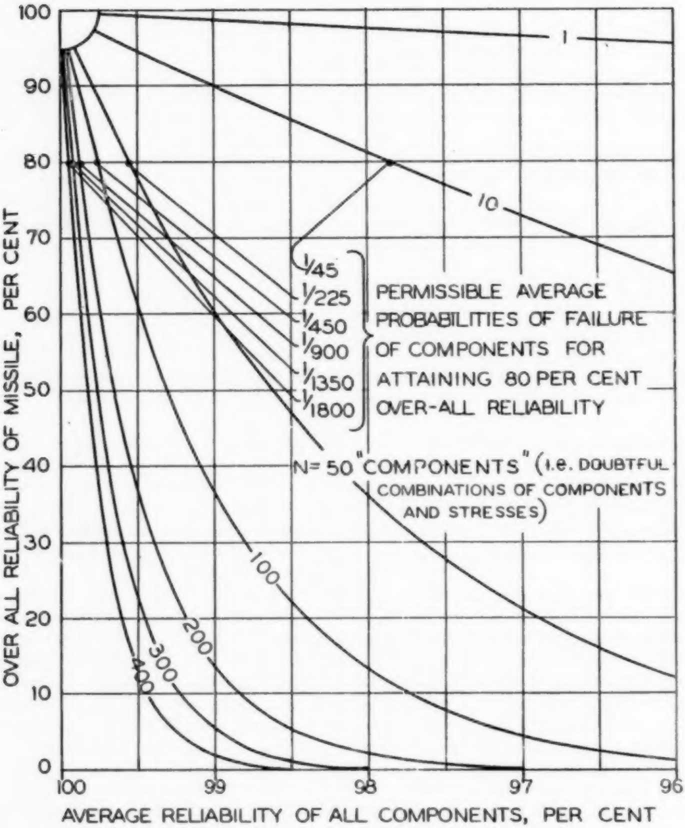
Let us translate this rule into engineering language: If two missile components have probabilities of success p_1 and p_2, the probability that both will function during the same firings is the product of the two probabilities: $P_{overall} = p_1 \cdot p_2$

Sometimes the applicability of this formula is questioned on the ground that it is not a perfect model for the rather complex relationship which exists between the reliability of the components and the overall reliability of the missile. This objection would be justified if an attempt were made to calculate a numerically accurate overall missile reliability on the basis of reliabilities of its components.

Such a computation, however, is not the intended purpose of the formula. Its value lies in the fact that it serves as a reminder to the designers, test engineers, manufacturers, and users of guided missiles, that the overall reliability of a missile depends on the product of the reliabilities of its components and by no means on the average. This fallacious notion regarding the average is still often encountered in the guided missile business. It may well be a predominant factor in the present over optimism which prevails with respect to the effort needed to achieve reliable and serviceable missiles.

The engineering significance of the relationship between the overall reliability and the required "level" of component reliability is profound. This can be seen by studying the diagram (Fig. 1, see following page) which is reprinted from NAMTC Technical Report No. 75.

Let us consider a missile with one hundred components, each having a reliability of 99 percent. By applying the $P_{overall}$ formula, we find that thismissile will have the amazingly low overall reliability of only 36.5 percent. This means that, on the average, two out of three missiles will fail. Similarly, if a missile had 400 components with this same 99 percent reliability level, 98 out of 100 missiles would fail!



OVER-ALL RELIABILITY AS A FUNCTION OF COMPLEXITY AND RELIABILITY OF COMPONENTS

FIG. 1

How reliable must these 400 components be made to insure a reasonable overall reliability of, say 80 percent? The $P_{overall}$ relationship indicates that a level of componet reliability must be attained so that only one out of 1800 units of each component will fail (see Fig. 1). This is a severe challenge for the designers and manufacturers of missile components.

One might well ask what is the presently accepted level of component reliability. An expert on statistical quality control recently said: "Let us say that we decide, for good engineering reasons, 1 percent defective is satisfactory. It is not unusual in a lot of 1 percent defective to get as many as three defectives in a sample of 100. Therefore, in order to avoid rejecting perfectly usable material, we have to allow three in a sample of 100."

This 99 percent standard has been satisfactory for those piloted aircraft components, numbering in the thousands, that are not really vital. For guided missiles, however, where all components are vital, this 99 percent standard is abolutely intolerable.

4. THE CHAIN ANALOGY

A missile is often compared to a chain that is just as strong, or weak, as its weakest link. This analogy is correct. It would, however, be a great mistake to believe that a missile is as reliable, or unreliable, as its least reliable component. In fact, it is far worse than that! As indicated by our reliability formula each component lowers the overall reliability by its own reliability factor, p. As a result the overall reliability is always much lower than the reliability of the least reliable component. For example, the least reliable component may have a reliability of .9, and yet the overall reliability might be .1, or less.

This fact leads to an important conclusion: Missile components having only a 90 percent, or even 99 percent reliability must be abhorred.

5. TRACING OF FAILURES TO THEIR ULTIMATE CAUSE

To remedy an unreliable component type it is necessary that the cause of a failure, such as poor welding of a structural part, or a short in a vacuum tube, be accurately determined. This is not enough, however. To prevent recurrence of the same failure in any subsequent unit, one must go farther and determine the ultimate cause of a weakness, such as:

a. Whether an environmental condition was more severe than had been anticipated.

b. Whether the specification of safety factors was too lenient.

c. Whether the component was poorly designed, or improperly selected.

d. Whether the component was poorly manufactured.

It is of particular importance to distinguish between design reliability and manufacturing reliability; these two categories pose problems

of an entirely different nature which must be solved by entirely different methods and by entirely different activities.

6. MANUFACTURING RELIABILITY

Let us discuss the manufacturing reliability first. A component may be designed perfectly yet fail because of poor manufacture. According to the basic reliability formula good manufacture by almost any other standard might be very poor manufacture for guided missile components. Ordinary standards, therefore, are not applicable. New yardsticks of quality and reliability must be developed and utilized.

Ordinary go, no-go inspection, even 100% or 200% inspection, is not enough. It is a "hindsight" measure at best, aimed only at the elimination of which has already been done wrong. And it usually does not help to eliminate the "last bug." However, it is the last bug that kills the missile.

By contrast, statistical quality control is a big step forward, since it is aimed at the prevention of future manufacturing errors. It controls and steadily improves the manufacturing process itself, and, therefore, attacks the source of trouble rather than its symptoms. Thus, it assumes an indispensable role in our struggle for attaining and maintaining the extreme level of component reliability required for guided missiles.

One of the chief advantages of statistical quality control is that it enables the manufacturer to strike a compromise between quality and cost which represents an economic optimum both to the consumer and to himself. To this end, a certain percent defective product is intentionally permitted. In ordinary applications this practice is quite satisfactory and acceptable. For guided missile manufacture, however, where each and every component, in the event of its failure, would "kill" an expensive missile, even a seemingly small percent defective is intolerable. A $100,000 missile may be a total loss because of the failure of a 10-cent component or part.

Therefore, in guided missiles we should never worry about making components "too reliable." Rather, we should strive for "absolute" component reliability, irrespective of cost. (The term "absolute reliability" is used here as a symbol for extremes in reliability, such as one failing unit out of ten thousand, or hundred thousand. There is, of course, no such thing as absolute reliability in a mathematical sense.) In guided missiles, a maximum reliability will, in the long run also mean maximum economy to the military services and the taxpayer. Therefore, the usual economic yardsticks of statistical quality control must be revised completely for components that are to be employed in guided missiles.

Thorough knowledge of statistical quality control is certainly essential to manufacturers and inspectors; a knowledge of at least its basic principles is desirable for designers as well if only to improve mutual understanding between designers and manufacturers. For further study see the following books:

W. A. Shewhart, "Economic Control of Quality of Manufactured Product." D. Van Nostrand, Inc.

Leslie E. Simon: "Engineers' Manual of Statistical Methods." John Wiley and Sons.

Eugene L. Grant, "Statistical Quality Control," McGraw-Hill.

7. DESIGN RELIABILITY

This phase of reliability poses a problem which is entirely different from that of manufacturing reliability and much more difficult to solve. A component, even when manufactured exactly as prescribed by the designer, may fail when subjected to the severe environmental conditions of launching and flight. Such a component is obviously inadequately designed. It is intrinsically weak. Unfortunately, there is little chance to determine the nature of these design weaknesses in flight tests or during service use, because missiles are not recoverable. An intrinsically weak component type might be pushed into mass production long before it has reached the high level of reliability required. Design reliability in the case of guided missiles, is a unique and critical problem.

Design weaknesses of a component can and must be traced to one of three categories of causes:

a. Environmental Conditions: Many environmental conditions are often only vaguely known, if they are known at all. Therefore, the actual conditions are frequently much more severe than those specified or than the designer has anticipated. A component can be very reliable when tested in an air-conditioned room without exposure to any shock, vibration, low or high temperature, etc., but in the climate of Alaska, or in front of a screaming rocket motor, the same component may fail immediately.

b. Strength: The strength of a component type varies from unit to unit with respect to a given environmental condition. This variation is usually much larger than the designer suspects, and is, therefore, often not considered in design. Even samples of a simple steel wire taken from the same bobbin may show a range in breaking strength which is 10 percent of the average strength. This well known fact is illustrated in Fig. 2a.

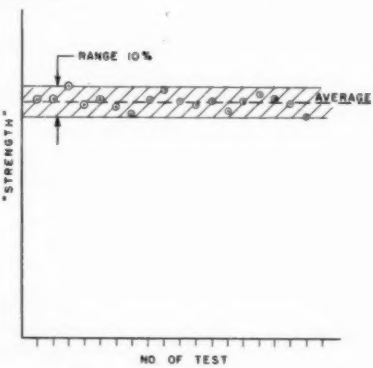


FIG. 2A

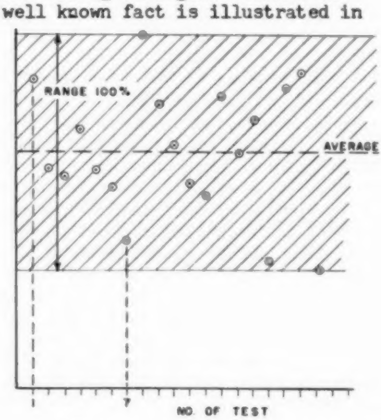


FIG. 2B

However, it is not well known that the strengths of the more complex and fragile component types may easily vary 50, 100, or even a greater percentage of the average strength value. This is illustrated in the theoretical example of Fig. 2b.

By comparing the great difference between test values 1 and 7, it becomes immediately clear that we cannot have faith in the result of a single test but must determine the characteristic variation of the test values as well. This is essential to making sure that even the weakest unit among thousands will not be weaker than the actual service stress.

It should be realized that just this inherent variation in strength from unit to unit, or rather the neglect of it, is often the real cause of missile failure. This variation cannot be calculated by the designer, nor can it be determined by flight testing and service use. There is only one way it can be determined satisfactorily and that is to test to failure a significant number of units of each component type. The strength data thus obtained will then enable the designer, the prime contractor, and the contracting agency to decide whether the safety factor attained complies with the specified safety factor. (The test to failure method is discussed at length in NAMTC Tech. Reports No. 75 and 84, and NAMTC Tech. Memo No. 70, which may be obtained from the Publications Department of the Bureau of Aeronautics, Department of the Navy, Washington, D. C., and also from Redstone Arsenal, Huntsville, Alabama.)

c. Safety Factors: Safety factors are often specified much too leniently. The safety factors presently specified for guided missiles were adopted from piloted aircraft where they are quite satisfactory.

You might well ask: Why are they satisfactory in piloted aircraft, yet not in guided missiles?

In piloted aircraft the structural safety factors are 1.15 with respect to yield and 1.5 with respect to ultimate (breaking) stresses. These moderate safety factors allow for only the natural variation in the strength of the basic materials such as steel, aluminum, etc. However, there is an additional and very powerful safeguard against breakage: the Design Load Factor. For example, a load factor of four is specified for commercial airplanes to protect against vertical loads caused by overcontrol and gusts. Thus the total safety factor specified against breakage (of a wing spar, for example) is $4 \cdot 1.5 = 6$. Now obviously the reliability of a wing spar depends not only on its absolute strength, but also on the maximum loads actually occurring in service. How high are these loads in the case of an airliner? Offhand one might presume that they are four times the total weight of the airliner, because this would be in accordance with the specified load factor of four. However, such is not the case. Commercial pilots are trained to control their ships so smoothly that they rarely exceed a load of more than about half again the acceleration of gravity. Even in very bumpy weather the pilot can prevent vertical loads which exceed 2g. (Incidentally, it is easy to verify gust loads of about 2g; when you are just slightly lifted from your seat, you have just been, or are about to be, subjected to a gust load of about 2g. You might try to remember whether you have ever experienced this. Possibly you never have.) Thus, even under extreme conditions of flight, we enjoy a comfortable safety factor of $6/2 = 3$.

In the design of guided missile components, particularly the nonstructural components, such as electronics, this design load factor is sadly neglected, apparently because it is felt that the internal components are not really vital. Actually, as was pointed out earlier they are not only quite as vital as the structural parts, but, because they are more numerous and more sensitive, they are also much more critical.

In contrast to the pilot of an aircraft, guided missiles do not sense the stresses to which they are subjected. They feel no pain and know no fear, and will, therefore, approach and exceed the many critical limits of their components without hesitation. This alone should justify the specification of high safety factors. Furthermore, because the number of vital components is extremely large, the level of component reliability must be extremely high. This would necessitate another considerable increase in the safety factors.

These facts, however, are not considered either in the specifications or in the design. Therefore, the safety factors of 1.15 and 1.5 are still the only specified safeguards against failure. To make things worse, the existence of even these much too lenient factors is seldom verified.

Unless new and adequate safety factors and margins are specified, included in design, and proved to exist, we will not have reliable guided missiles.

8. FOUR BASIC PRINCIPLES

To help overcome the present inertia in the guided missile reliability situation, the following basic principles are offered.

a. The presently specified environmental conditions, limiting test values, and safety factors, cannot be trusted.

b. The actual environmental conditions of service, and their variation, must be carefully determined by scientific tests.

c. Unusually high safety factors must be specified and applied, to insure virtually "absolute" design reliability of all components.

d. The existence of these high safety factors must be proved by testing significant numbers of units to failure.

9. THE IMPORTANCE OF STATISTICAL CONCEPTS

It is significant to note that these rules are strongly intertwined with the laws of probability and statistics. What does it mean to specify that a component must be "very" reliable, or that "maximum" reliability must be achieved? Such nonobligatory terms mean nothing.

Would you think, for example, that a failure rate of 1:1,000 is satisfactory for a certain type of vacuum tube? It might be -- if a missile contains just one or two of the type. However, if it contains 200 such vacuum tubes, this failure rate of 1:1,000 would be catastrophic, because about one missile in every six could be expected to fail just because of this type of vacuum tube.

One more example: Some time ago I visited a manufacturer of elec-

tronic harnesses for a guided missile. The inspector didn't know what his rate of defective soldered joints was, but readily agreed that he would be well pleased with a rate of one undetectable cold soldered joint out of 5,000. When I explained to him that this particular missile contained about 5,000 soldered joints, and that, therefore, this failure rate of 1:5,000 would cause the failure of about 63 percent of all missiles, he was startled. He complained that no one had ever told him that with these particular soldered joints he should strive for a rate of defectives of, say 1:100,000, or better, 1:1,000,000, and expressed a strong desire to learn more about reliability control.

The inspector cannot be blamed for his failure to achieve the required extremely high level of reliability. It is the designer who must figure out what level of component reliability must be striven for in each type of component. It is he who should personally go to the inspector and tell him why this extreme level is imperative, how it may be achieved, and how it should be checked and proved.

Here I should emphasize one very important aspect of design reliability: It is not just the few environmental conditions, such as shock and vibration, which deserve our attention. There are hundreds and possibly thousands of design considerations which might be critical. Each component type has at least some functional design criteria that may be hazardous. In each case the designer must strive for "absolute" design reliability so that a malfunction is not likely to occur in more than one unit out of ten thousand, and in the case of very complex missiles, even more. For example, a container should never burst or leak; the servo oil should never be consumed before intercept, or impact; the torque of a servo should always be strong enough to activate the rudder; the thrust line of a rocket should always be aligned through the center of gravity; and no carrier frequency should ever deviate beyond its tolerance limits.

It should be realized that even doubling or trebling the intrinsic reliability of a component type is a real challenge to the designer. Imagine how much more severe is the challenge of striving for a level of component reliability that is a hundredfold better. The basic laws of probability and statistics are essential to such an achievement.

10. ROLE OF THE DESIGNER

The question arises: Just who must know probability and statistics? Is it the inspectors? The statistical quality control specialists? The reliability coordinators? Yes, for those it is certainly a must. However, it is the designer who carries the main burden of responsibility for the reliability of his brain child. No one knows better than he the critical design criteria or weak spots of his component. No one can have a keener interest in tracing weaknesses and failures to their ultimate causes and finding the proper remedies. The responsibility of the designer cannot be relieved in the least by statisticians, specification writers, inspectors, checkouts, or so called preflight "environmental screening."

Thus it is imperative that all of the many thousands of designers concerned with a missile and its components understand the basic concepts of probability and statistics and their application.

One should not approach probability and statistics with the idea

that it is sufficient to hear an occasional talk on the subject, or to read an article addressed to the layman in a popular magazine. It really has to be studied. In fact, at first the conquest of statistics may require some self conquest as well, but as soon as the common aversion to it has been overcome this important field of science is generally received with a great deal of enthusiasm.

Unfortunately, few of you will be able to spend as much time on this study as you might like. This must not discourage you. You do not need to become full-fledged statisticians -- far from it. The reliability problems of your components can, and must, always be solved by mostly engineering judgment plus a comparatively small amount of statistics -- and by no means the reverse!

You might want a suggestion on how to get the necessary elementary knowledge in probability and statistics quickly and efficiently. Well, a great variety of books is available to you. Some are easy to read, but many others are so highbrow that you might immediately become discouraged; beware of these.

Excellent textbooks are Waugh's Elements of Statistical Method, published by McGraw-Hill, and Freund's Modern Elementary Statistics, published by Prentice-Hall, Inc. You will find them very easy to read. The information you need for your work is contained in the first half of the books.

As soon as you are through with this elementary study you will feel very happy that you have acquired a new habit of thinking, because virtually every problem of life can be clarified by some knowledge of probability and statistics. From then on you will enjoy reading special studies in reliability, and soon you will be a strong link in the long chain of designers, test engineers and manufacturing specialists concerned with guided missiles and their components. Since this chain, too, is no stronger than its weakest link, no designer, no test engineer, no production engineer should exempt himself from this obligation.

11. THE TEST-TO-FAILURE METHOD

To find the proper remedy for a component type that has caused the failure of a missile, it is imperative that the failure be traced to one of the four ultimate causes: (1) poor design; (2) poor manufacture; (3) poor knowledge of an environmental condition; (4) specification of safety factors or test procedures which are too lenient.

In guided missiles, this tracing to the ultimate cause cannot be done by flight testing, because missiles are not available for post-flight inspection. One of the most powerful means of implementing such a development is the systematic laboratory testing of all component types, up to failure, in statistically significant numbers, with increasing severities of environmental conditions and design criteria. This is called the test to failure method. It is discussed at length in NAMTC Tech . Reports No. 75 and 84, and in NAMTC Tech. Memo. No. 70.

The data of ultimate strengths of a component type obtained by tests to failure can easily be plotted as shown in Fig. 3.

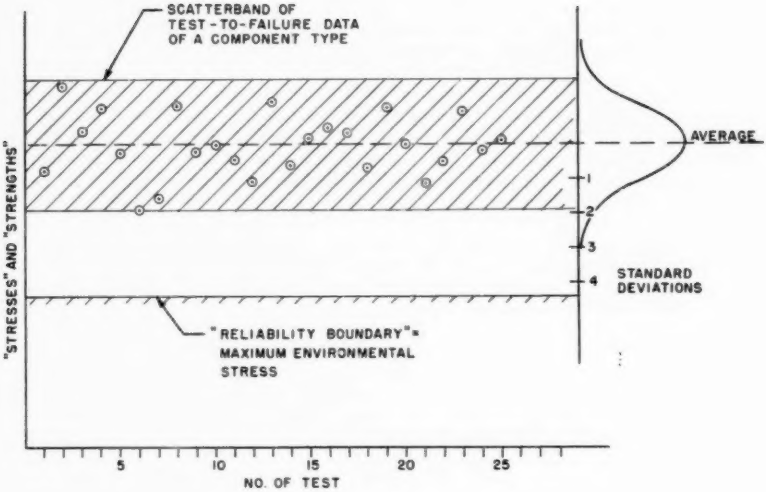


Fig. 3

As mentioned before, the width of these scatterbands varies enormously, depending on the type of component and the environmental condition. It is for this reason that the designer of a component must know, by all means, not only the "average strength," but the characteristic width of the scatterband, expressed in standard deviations. Only then will he be in a position to judge whether a component type is strong or weak, consistent or inconsistent, reliable or unreliable, with regard to a particular stress condition encountered in service. There is virtually no other way to find this out.

Furthermore, the test to failure method will provide a rational basis for specifying that a certain minimum number of standard deviations, say five, shall be available between the average value and the so-called "reliability boundary" (the specified maximum value of an environmental condition).

12. ECONOMIC ASPECTS OF TEST TO FAILURE METHOD

There is a great economic advantage inherent in the test to failure method. The number of tests to failure required for proving the strength and reliability of a component type is very small compared to the huge number of tests required when the reliability of a component type is to be obtained from tests under the severity of actual flight conditions. In the former instance we might need five, ten, or twenty tests, in the latter, possibly several thousand. (See NAMTC Tech. Report No. 75, pages 26 - 29.)

The most important advantage of the test to failure method, however,

lies in the fact that we can study the defects and malfunctions at close range, and trace them to their ultimate causes. This will, in most instances, lead to quick, accurate, remedies and improvements.

It may be argued that testing to failure might be too expensive and time consuming. This criticism does not appear justified, however, when test to failure cost is compared to that of the enormous monetary and military consequences which result from inadequate component reliability. We must always remember that even one seemingly unimportant component type can ruin a whole missile weapon.

Test to failure programs can be planned and conducted very economically and efficiently. No component type should be omitted from such tests to failure, even though it seems to be strong and reliable. However, if the first unit so tested turns out to be, say four times stronger than the maximum condition encountered in service, it is generally not necessary to test any more units. In most instances a safety factor of four usually eliminates the danger that any subsequent unit may be weaker than the maximum service condition.

Many component types will never be four times stronger than the maximum service condition. This is particularly true of structural parts where saving of dead weight is an important factor. However, in all instances where the first unit reveals a safety factor of less than four, one must determine the natural variation of the strength from unit to unit. For this purpose, more units must be tested to failure until statistical proof is obtained that a specified safety margin of at least five standard deviations is attained. Then and only then can we hope that the necessary very high degree of component reliability has been reached. (For further discussion, see NAMTC Tech. Report No. 84.)

13. WHEN SHOULD A TEST TO FAILURE PROGRAM BE INITIATED

A comprehensive reliability test program, aimed at the detection and elimination of weakness long before the first missile is fired, is essential to any well-managed missile development. Such a program should result in a rapid rise of the reliability level of the many components, and consequently the weapon may reach a serviceable state much earlier.

The first phase of such a program should deal with all of the hundreds of standardized components and parts, such as valves, relays, and vacuum tubes. These are the real building stones -- and hazards -- of a missile.

Certainly these items should not simply be selected from catalogues without any knowledge of what conditions they can really withstand, and without knowing whether or not adequate safety factors and margins exist. They may be reliable enough for ordinary applications in piloted aircraft, radio, and television; there they are not vital, yet not at all appropriate for use in guided missiles. If they are not intrinsically highly reliable and perfectly manufactured, a missile weapon may never reach a truly reliable, serviceable state.

There is a general hope and belief that these components can be made reliable at a later stage of development. This hope is vain, however, because the design of the components must be frozen once the order to start mass production is given. Weaknesses of component types might, therefore be caught in the inexorable mass production process and constitute severe

permanent hazards to the missile weapon.

<u>Therefore, to be of maximum benefit, a reliability test program should be started at the very beginning of missile development and conducted under highest priority so long as the missile is being produced.</u>

A comprehensive reliability test program for a guided missile and its hundreds of component types is not a quick and easy job. It might be a matter of tens of millions of dollars, rather than tens of thousands.

The savings which can be realized by fully utilizing the fundamental engineering principle of testing to failure will far outweigh its cost, however. This is because it will enable us to build a firm foundation before we put the roof on, so to speak, and help eliminate the usual patchwork, which will not lead to reliability anyway.

PRACTICAL EXPERIMENTAL DESIGNS IN CHEMICAL RESEARCH

Donald S. McArthur
Esso Research and Engineering Company

Research is done to obtain information. An industrial organization must obtain information about the laws of nature which pertain to its business, if it is to prosper and be of optimum service to society. Each unit of information has two parts; a conclusion and an estimate of the risk involved in reaching the conclusion. The usual way of doing research, trying to study one variable at a time, often is not the most efficient. Designed experiments produce more information per year, they reach more conclusions and give better estimates of the risks.

In experimental research, we seek information by looking for relationships between variables. Our problem is to learn how certain independent variables X_1, X_2, X_3, --- affect some dependent variable Y. The general problem can be expressed as one of determining the function in the relationship:

$$Y=f(X_1, X_2, X_3, ---)$$

For example, we might want to know how oil additive type (X_1), fuel type (X_2) or some engine operating condition (X_3) affects piston ring wear (Y). The function will not usually be represented in the form of a mathematical equation but by graphs or simply by words (e.g. oil additive K reduces engine wear).

The expression above represents an oversimplified research problem however. We are usually plagued by a host of unwanted or uncontrolled (U type) variables which can have a big effect on the dependent Y variable in which we are interested. Most research problems might be expressed better as follows:

$$Y=f(X_1, X_2, X_3, --- U_1, U_2, U_3, ---)$$

In an engine wear program, for example, U_1 might be air humidity, U_2 might be differences between "identical" engines, and U_3 might be engine operator, and so on. Uncontrolled variations in the U variables cloud the relationships between the X variables and Y.

THE CONVENTIONAL WAY OF DOING RESEARCH

The usual way of doing research is to vary one X variable at a time while trying to hold the others constant. This way of doing research has two weaknesses. The first is that it is apt to miss interaction effects between the X variables. The effect of variable X_1 on Y may depend on the level of X_2 and X_3. For example, the effect of oil additives on engine wear may be different for one fuel (or operating condition) than it is for another. The effect of the different X variables (on Y) may not be simply additive.

The second weakness of the usual way of doing research is that it makes it difficult to take enough account of the uncontrolled variables. Y may be affected by both the X and the U type variables. When a change is observed in Y, a decision must be made; was that change caused by controlled changes in the X variables or by uncontrolled changes in the

U variables? In doing research the usual way, this decision is left to our experience or judgment.

DESIGNING EXPERIMENTS CAN HELP

The two weaknesses in orthodox research can be overcome by the use of statistically designed experiments. The designed experiment helps us detect interaction effects between the variables by suggesting a different way of doing research. Rather than varying one X variable at a time, the designed experiment varies all of the X variables at the same time. Doing this in an organized way gives us the effect of each X variable over a range of each of the other X variables. If interaction effects exist, we can detect them and get a quantitative estimate of their importance.

The designed experiment takes the U type variables into account. Statistics takes advantage of the difference between the way the X variables and the U variables change. The X variables change in an organized way while the U variables change in a disorganized way. Upon the completion of a designed experiment, the statistician can take advantage of this difference to calculate the odds on the situation. For example, he might tell us that there are 9 chances in 10 that the reduction observed in engine wear was due to the use of additive K in the oil and only 1 chance in 10 that it is due to a change in one or more U variables. The designed experiment gives us better "resolving power" than the old way of doing research. We can see effects more clearly through the inevidable fog of experimental error. The design of experiments can't replace judgment in research work but it can improve our judgment.

HOW TO DESIGN AN EXPERIMENT

Six steps seem to be common to the design of most industrial experiments.

1. The first step is to list the variables. What do we plan to measure? What is the Y variable? (There may be several.) What factors (X variables) are we interested in? What are the uncontrolled U variables likely to be? Every bit of chemical or engineering knowledge of the problem will be needed in designing a good experiment.

2. Then we should consider what we already know about the system. Are the X variables, listed in step 1, simply guesses or do we know from past work that they have a real effect on the dependent variable? There are a large number of experimental designs available. The one we choose will depend on the problem we have before us. If they are simply guesses, we want a cheap experiment which will sift out the unimportant variables. If we know that they are important, we want an experiment which will tell us in more detail what their effect is.

We should consider whether we have an estimate of the error in the test from past experience. If we do, we can make use of this information in designing the test.

3. We must decide what resolving power is needed in the experiment. What is the minimum change in Y which is of practical importance? There is no point in designing an experiment with a resolving power sufficient

to detect differences of 0.1% in Y if a difference of less than 1.0% is of no practical importance. On the other hand we must try to design the experiment so that it will detect differences of 0.1% if that difference is important.

4. We should consider whether interactions between the variables are likely to be important. This will affect the type of experimental design used.

5. Then we should list the practical limitations in running the experiment. Are we limited by time or by equipment? Are there peculiarities in the test which must be considered?

6. The final step is to design the experiment. The experiment used will come as close as possible to producing the information desired while staying within the practical limitations noted above.

FOR EXAMPLE

In the Esso Research and Engineering Company we recently designed an experiment to give us more information on engine wear. In designing the test we went through the six steps listed above.

1. What Are the Variables?

The experiment was designed to study piston ring wear in an engine. The dependent variable (Y) was ring wear. Two wear rates were of interest; start-up wear (the wear occuring during the first 15 minutes of engine operation) and running wear (the wear occuring after this 15 minute period). There were two dependent variables; start-up wear Y_1 and running wear Y_2.

It was known from previous experience that engine wear can be affected by the fuel and the lubricant used. It was desired to study these variables more closely. In particular, we wished to learn the effect of three different fuels, A, B and C and to compare a base oil with the base containing three oil additives K, L and M. We also wished to learn whether the length of time the engine had been shut down prior to the test had an effect on engine wear. Do we get more wear in the engine when it is started after a weekend shutdown than when it is started after an overnight shutdown?

2. What Do We Know Already?

Previous experience had shown us that fuel and lubricant variables can affect engine wear and that shutdown time can affect engine wear. We were interested in studying the effects more closely. The previous work had also shown that the error in the engine wear test is large (the U variables have a big effect on wear).

3. What "Resolving Power" is Needed?

In order to detect a difference of practical importance we needed at least 6 engine tests in each comparison.

4. Are Interactions Important?

Yes. It seemed quite possible that interactions between the fuel and the oil additives would exist. We were definitely interested in interactions between the oil additives and the shutdown time. An oil additive which would nullify the effect of shutdown time would be of real interest.

5. What are the Practical Limitations?

There were several practical limitations.

(i) Only one test could be run per day in each engine. The prior shutdown period was part of the test. Each test was started at the end of the working day by flushing the engine out thoroughly and then running it on the test fuel and oil. The engine was then shutdown for the night. Next morning it was started up for a 3 hour run. A wear measurement (Y_1) was made at the end of the first 15 minutes of running and another one ($Y_1 + Y_2$) was made at the end of the 3 hour test. The wear occuring during the running period could be determined by difference. The engine wear tests were to be run using a radioactive top piston ring. Measurements of the wear were made by determining the amount of wear debris (radioactivity) in the oil.

(ii) It was not practical to run the engine (without overhaul) for more than one month. It was known from previous experience that overhaul in the middle of a test program disrupted the program badly. The whole experiment must therefore be completed in one month.

(iii) Only two test engines were available.

(iv) The wear levels in the two engines might differ.

(v) The lubricant additives were such that there was a danger of hangover effects from one test to the next. The additive used in yesterday's test might affect today's test in spite of careful flushing of the engine between tests.

(vi) There are only 4 weekends in a month. It was not practical to design a program with as many weekend shutdowns as overnight shutdowns in it.

(vii) There was a good possibility that the weather would change during a month-long program. It was desirable to minimize the effect of the weather (and other U variables) on the program.

HOW THE PROGRAM MIGHT HAVE BEEN RUN USING THE ORTHODOX APPROACH

Two test engines were available. Experience has shown that the results from two engines may not be interchangeable because of (unknown) differences between them. Using the orthodox approach we might have decided therefore to study oil variables in one engine and fuel variables in the other. The study of oil variables would be carried out using a standard fuel and a fixed set of engine operating conditions. The different fuels would be studied using a standard oil and fixed operating conditions.

There is usually considerable doubt about how many tests should be run on each oil or fuel to show significant differences. This might lead to running only one or two tests on each oil and fuel and drawing (erroneous) conclusions from these results. Let's assume that experience leads us to run six tests (the proper number) on each oil and fuel. The three fuels would require 18 tests in engine number 1. This will consume one month of operation on this engine. The Monday morning tests can be used to give an idea of the effect of shutdown time on wear. Engine number 2 can be used to study 3 oils (the base oil and additives K and L). In the ordinary case the fuels and oils are tested one after the other as ideas occur to the experimenter. This doesn't permit scrambling tests to minimize the effect of variations in the weather (for example).

Both engines must now be overhauled. Experience has shown that overhauling an engine can have a drastic (and unpredictable) effect on its wear rate. This change in wear rate must be determined experimentally if tests after the overhaul are to be compared with tests run before overhaul. About 6 tests on a reference oil and fuel will be required to establish the wear level of each overhauled engine (12 tests) with the reliability needed to compare oils. One engine can then be used to run the fourth oil containing additive M (6 tests). The other engine could be used to investigate possible interactions between the oils and fuels. If 2 tests are used on each fuel-oil combination, this part of the program will require over 20 additional tests.

The entire program run in the orthodox way would require over 60 engine tests and more than two months of test time. The same information can be obtained in half the time using a statistically designed experiment.

HOW THE PROGRAM WAS RUN

The design of an experiment depends upon the ingenuity of the experimenter. Two broad types of experimental design are available, the Factorial and the Latin Square type. Many modifications of each type have been used. We have found the factorial experiment to be the most useful. A small fraction of a 2^n factorial test is useful in screening variables for importance when we start simply with guesses. The full factorial gives us a good understanding of how the important variables affect the problem and of how they interact with each other. This experiment was designed as shown in Table I. A total of only 32 tests were run. The design gave information on the effect of the three X variables (fuel, oil and shutdown time) on the two dependent variables (start-up wear Y_1 and running wear Y_2). The results were better than would have been obtained by the conventional approach.

The experiment was designed so that at least 6 tests were averaged in every comparison. This gave it the resolving power necessary to pick out differences of practical importance in spite of the large error inherent in the test. Interaction effects could be detected. The interaction between fuel and lubricant additives could be determined. The interaction between oil additive K (the additive of most interest) and shutdown period could be estimated. However, any interaction between the other oil additives and shutdown time could not be obtained.

All of the tests were completed within one month's time using the two test engines available. The design of the test was such that a

difference between the wear levels in the two engines did not affect the test results. The effect of changes in the weather (and any other U variables) was minimized by the choice of the test sequence as shown in Table I. For example, all the tests run on any one additive were not made in one week. Possible hangover effects from one test to the next were minimized by running the tests in one engine in the reverse order to that used in the other engine.

These wear results must of course be checked in the field. Field engines may respond differently than laboratory engines to the variables studied. However, the test program came close to providing all of the desired information obtainable within the practical limitations imposed by the available facilities and the wear test.

WHAT INFORMATION WAS OBTAINED?

It was discovered that gasolines B and C gave significantly higher wear than gasoline A. Their effect was only apparent during the running period, not during the start-up period. The lubricant additive K reduced engine wear significantly. Neither L nor M affected the wear. It was discovered that there was a significant interaction between fuel B and lubricant additive K. Although K was an anti-wear agent it reacted with fuel B to form a prowear agent. Longer shutdown periods increased engine start-up wear but had no effect on the running wear. Lubricant additive K did not nullify the effect of shutdown time. It was discovered that engine number 2 gave significantly more start-up wear than engine number 1. They were equivalent in running wear.

Each positive conclusion reached above had a calculated risk attached to it. All conclusions had at least a 90% chance of being real ones. In most cases there were less than 5 chances in 100 that the observed effect on wear was due to one or more of the U type variables.

SUMMARY

Research work is done to obtain information. A unit of information has two parts; a number or conclusion and an estimate of its reliability. In research we obtain information by investigating the effect of several independent (X) variables on one or more dependent (Y) variables. However, relationships are clouded by the presence of uncontrolled (U) variables. The design of experiments can help us get more units of information per year. It accomplishes this by increasing the resolving power of the experiment, by reporting interaction effects as well as main effects between variables, and by giving us a calculated risk to assist our judgment in using the data.

The best kind of experimental design depends upon the problem. The factorial design is often useful in petroleum research. The fractional replicate of the factorial experiment can be used to pick out the important variables from the unimportant ones. When the important variables have been established, the full factorial test will give an optimum amount of more detailed information on them. An example of the use of a factorial test in engine wear studies demonstrates its usefulness.

TABLE I

DESIGNED TEST ON WEAR

LUBRICANT ADDITIVE:	None		K		L		M	
TESTS RUN AFTER AN OVERNIGHT SHUTDOWN								
Engine No.	1	2	1	2	1	2	1	2
FUEL								
A	1(1)	12	6	7	12	1	7	6
B	9	4	2	10	4	9	10	3
C	8	5	11	2	5	8	2	11
TESTS RUN AFTER A WEEKEND SHUTDOWN								
A	1M(2)	4M	3M	2M				
B	-	-	-	-				
C	4M	1M	2M	3M				

(1) Numbers in the table show the test sequence in that engine.

(2) The M stands for Monday. These tests were run after the weekend.

INDUSTRIAL ENGINEERING AND QUALITY CONTROL .....
..... MANAGEMENT'S ANSWER TO THE COST PROBLEM

George F. Bluth
Studebaker-Packard Corporation
Detroit, Michigan

Throughout the world, in practically every type of business, men have devised slogans and "catchy" phrases as part of their master plan to merchandise their products. Such phrases as:

..... "You can be sure if it's Westinghouse" .....
..... "Philco stands for Quality the world over" .....
..... "You can place your confidence in General Electric" .....

are only a few. This philosophy, however, represents the fact that Quality is recognized as one of the most conclusive features of a product, upon which customer acceptance and profit may be realized.

For many years, Industrial Management has realized the fact that Quality, coupled with manufacturing cost are the two variables in their profit picture which contribute primarily to the success of a business enterprise. Because of this, they have utilized many of the modern day applications of automation, cybernetics, time and motion study, cost control, and others, in an attempt to derive consistant and economical manufacturing processes, upon which Quality programs and cost control techniques could intelligently be formulated.

These expensive and technical applications of manufacturing methods and facilities have saved many a dollar for modern-day industry. Management has also realized, however, that the savings obtained by using these techniques can often be offset by having large quantities of the products which are produced by these new methods wind up on the scrap or rework bench.

As a result, Industrial Engineering, with its many facets, along with Quality Control, have been assigned the task of keeping our methods and techniques abreast of our facilities and market conditions.

With the ever-improving methods of modern industry, both the Industrial Engineers and the Quality Control people have discovered that the requests for their services have increased to the point where there is hardly time to completely analyze a situation, due to the complexity of operations and the terrific pace of industry today.

For this reason, it becomes evident that in most cases, many Industrial Engineering and Quality Control programs have not kept pace with other improvements in the fields of styling, plant engineering, research and sales. Programs for the control of quality and cost must be based now, more than any time previous, on utilizing the most simplified and expedient techniques to arrive at a point of preventative control rather than corrective action.

Management, today, must take advantage of improved inspection techniques and special tools which are now available to handle the complexities and problems of our manufacturing processes. We must look forward to the demands that our economic future is presenting and recognize the aid that young, specially-trained technical men can offer.

Sometimes in the heat and haste of mass-production these facts are overlooked.

Actually, the science of Industrial Engineering is not new to the manufacturing business. Like Quality Control, Industrial Engineering has spent many years of constructive planning and research to come up with an intensified plan which would help to improve manufacturing operations, by pointing out economic inefficiencies and establishing various programs of Methods and standards, which would allow production personnel to run their business in an orderly and economical fashion.

In the meantime, Quality Control has approached the same situation by deriving many integrated systems of statistical techniques, designed to provide an intelligent basis for gathering the type of information which points out areas for cost improvement.

For many years, Industrial Engineering and Quality Control have each traveled in their own direction, accomplishing what time and manpower could afford, without realizing that the net result of their efforts were essentially the same. As a result, many organizations have realized that by exchanging and correlating the knowledge and efforts of both activities, both the Industrial Engineers and the Quality Control people are in a better position to accomplish their objectives, by realizing and using the techniques of each science. In this fashion, both organizations can accomplish their tasks and utilize their own manpower to keep pace with industrial conditions, thereby fulfilling their responsibilities of establishing preventative control programs.

When we speak of the science called "Quality Control", we are speaking of one thing only ..... the control of quality ..... Quality, just like anything else, in order to be controlled must consist of the following elements.

I. We must know what kind we are talking about. We must have a definite standard of quality, that is made known to all of us who must live by it.

II. We must have the services of specialized people who are in a position to ascertain the causes of our Quality problems, and direct the necessary steps to resolve them.

III. We must have a continuous program of quality improvement, based on preventative control, rather than "locking the barn after the horse is stolen" ..... or ..... "fixing the job after it is out the back door."

IV. We must have sound inspection methods; we must have good tools; and we must have the best of gauges that are capable of enforcing quality standards that we establish, with a maximum of precision and minimum of personnel.

V. We must have a common understanding between ourselves, our vendors, and our service personnel, so that all of us collectively are driving in the same direction ..... and,

VI. We must have a program of analysis whereby the results of a system may be measured.

For many years, the Quality problem was faced by <u>many</u> organizations by developing a hard-hitting inspection force, which would advise production whenever they did the wrong thing. This proved to be a very costly and nerve-wracking policy because, as industry found out, quality cannot be inspected into a product.

At this time, a new science, like Industrial Engineering, was striving to show management a new approach to their problem. This science was called Statistical Quality Control ..... which simply meant the control of quality through statistical methods.

Many organizations today have established progressive and highly technical Statistical Quality Control programs. The approach used in establishing these systems varies considerably from one organization to another, as well as from one type of industry to another. A typical Quality Control organization, however, usually consists of two individual components ..... an Inspection Department and a Quality Analysis Department, each with its own authority and responsibilities, but collectively providing a program for the control of Quality. This fundamental is essential. A Quality Control program <u>must</u> actually control the quality ..... otherwise it is not effective.

Most Inspection Departments consist of the many technically-trained and seasoned inspectors, ranging from lay-out men and receiving inspectors to final line and floor inspectors, whose primary job is to inspect both the product and the areas in which it is made ..... and to record the results of their inspection.

A Quality Analysis Department consists likewise of a group of technical engineers and statisticians, whose function is to coordinate the results of inspection in such a fashion that the specific cause of poor quality conditions may be ascertained. In addition, they completely analyze the quality picture by compiling statistical evaluations of every phase of the company's business which pertains to quality.

The collective efforts of these two organizations provides a sound plan for disseminating the type of information and recommendations that are constructive to the well-being of a business enterprise. Such a program, in itself, has proved to be a tremendous weapon for cost control, but it is not enough today to rest on any laurels.

If we, in the Quality business, are to keep pace with our partners in industry, we must provide for production and administrative management, a program of analysis and control so complete that there is no room for doubt as to the causes of our problems ..... or room for excuses, where preventative action is needed.

In addition to statistical charts and procedures, our quality control programs must consist of evaluations and cost studies, based on the capabilities and limitations of materials, machines and manpower. Our methods and statistical applications must be flexible enough so that our programs may be readily-adaptable to whatever changes may take place in our production operations ..... such as the acquisition of automatic equipment or the installation of more efficient gages and tools.

Our analyses and efforts should be tailored to accurately prove the evaluation of new equipment and methods, so that positive cost

studies and decisions can be determined before a company has invested too much time and money. This is essential, since expenditures and savings of this type usually affect the market price of the company's products.

In considering the three elements of quality ..... materials, machines and manpower, it becomes apparent that certain factors exist which, under normal circumstances, are serious problems, but under the influences of cost-saving programs and automation, become primary, or chronic, conditions.

Positive Quality standards and inspection standards must be established throughout every operation, so that manpower or productivity will not be lost because of differences in people or judgement.

In Studebaker-Packard's program, for instance, Inspection Instruction Sheets have been provided for each inspector in every area, which clearly spell out <u>what</u> he is to inspect, <u>how</u> he is to inspect it and <u>how often</u> the inspection must be made. This provides not only for the proper classification of defects, but enables certain attributes or variables to be measured on the particular type of statistical plan which is best suited for that particular case. This one technique, in itself, has enabled the proper allocation of inspection manpower to be made and places in the hands of the inspection foreman, or supervisor, a tremendous aid for training his people quickly and properly. This feature makes possible a sound evaluation of work load and permits better accuracy in estimating and measuring certain indirect labor costs.

Under this plan, inspection manpower can then be utilized in the areas where this type of labor can show the best results. As we are well aware of, inspection efforts can justify themselves best when quality and cost problems are kept within the areas that manufacture them. For this reason, inspection now becomes a positive link in any control system.

The materials program, or receiving inspection, as it may be called, consists likewise of a positive charge-back and evaluation program, in addition to the regular techniques of sampling plans, instruction sheets and vendor performance. Defective material and the direct or indirect labor that is utilized to screen, sort or repair such material, are paid for by the contracting vendors through a positive evaluation of each part received and a system of bookkeeping which makes it expensive for a vendor to default in his agreements. Industrial Engineers have found that this system allows them to intelligently forecast and measure standard and non-standard costs, which previously could have been hidden under the general classification of overhead or burden.

Since machines and the tools which are used in machines are commonly the cause of many quality problems, as well as cost increases, our control programs in these areas are based primarily on process and machine capabilities with an accent on tool life evaluation.

Process capabilities are conducted first, to locate the areas where machine or operator capability studies are needed, so that our efforts will be used wisely and in locations where the most benefit can be obtained.

Too many organizations have applied the technical statistics of

capability tests in areas only where manpower or conditions warranted it. This mistake often nullifies the benefits which this form of evaluation can render.

In most of our machining areas, a complete tool life evaluation program has been established on the expected usage and actual results of every tool used in every machine.

Toolometer boards are used in most areas, particularly along automation lines, which make possible the accurate measurement of tool usage. These boards are devised in such a manner that each tool in every machine is controlled by the number of pieces which it produces. As pieces are produced off of any given machine, a dial on the toolometer board registers electrically the actual quantity produced against the estimate which was established originally. When a jobsetter replaces a tool at the end of its toolometer run, or sooner, if necessary, a pre-coded card is filled out, indicating the actual pieces produced by this tool and the reason why the tool was changed. This card accompanies the used tool to the cutter grinding department, where the height or length before and after grind are recorded. The reground tool is placed back in the toolometer board and the card is key-punched and historically analyzed.

Through this system, the usage of each tool and each type of tool can be based on sound, statistical life expectancies. The ultimate in tool life is realized from each tool, yet the most economical conditions are accomplished by eliminating unnecessary downtime.

In addition, this system enables purchasing activities to establish regular and economical methods of ordering new tools. Tool inventories can then be controlled wisely and the problems created through excesses or waste are eliminated.

By controlling the tool part of our machining operations, machine capability studies can then be narrowed down to the point of common sense, resulting usually in operator or shift variances.

Effectively controlling the materials, machines and tools, our quality control personnel can now become an active part of the day-today production problems, rather than applying mathematical formulas where convenient, in a general attempt to save money.

Along the same line, a preventative maintenance program on welding equipment and electrode dressing has proven invaluable in obtaining consistant welding practices and elimination of unnecessary downtime.

This preventative maintenance plan provides essentially the same benefits as a tool life evaluation, establishing a definite system for tip-dressing and gun repair. Cost studies are then based on a stable pattern, since an accurate estimate can be determined on any new operation, utilizing the experience which was acquired previously. Up until this time, most of the welding experience of any company has passed out the front door with every retired worker.

This plan, or any other type of controlled maintenance plan, can only be effective, however, <u>after</u> definite quality standards have been established. Without a common understanding and agreement of quality standards, programs which enforce those standards are futile.

Special capability studies on production operators and shifts also help to point out quickly where extra cost and unbalanced conditions exist, as well as to indicate the results of previous training programs or newly-installed methods. The accuracy of these studies, of course, depends largely on the adaptability and sensitivity of the analysis techniques employed.

Previously, operator capability studies were conducted by many companies only when a known difference existed. With this new type of approach, an operator capability study is conducted as a regular part of the plan, since the materials, machines and tools are measured and controlled separately. Through a quick and accurate process of elimination, a problem can usually be resolved before an operation becomes a crisis.

In addition to the ordinary methods of control charting and reporting of defects, Studebaker-Packard's quality program consists of a complete, detailed analysis of field complaints by type of defect, geographic area and frequency. These key-punched reports are associated by date of production, with all in-process records concerning the fabrication and assembly of the particular unit in question.

From each tool and machine, through every engine, transmission, chassis and body, an integrated system of key-punched cards records the quality efforts applied to each finished automobile. In this fashion, the cause and responsibility of any field complaint can be ascertained and charged to the appropriate area or person concerned.

The net result with this type of control program, provides a workable system which can be understood by all concerned, since it incorporates the active participation of every person from a jobsetter or inspector up to the vice-present level.

All of the information and benefits which a system such as this provides is only as valuable as the attention which is given to it. For this reason, any reports or recommendations which are submitted to management include not only the causes of quality or cost problems, but spell out precisely, with the help of the Industrial Engineers, the excess costs which develop as a result of any sub-standard condition incurred by any plant or department. These costs are summarized, taking into consideration every facet of information available and presented to plant management in the following fashion:

> Total amount of excess cost is divided by the total number of units which are produced during the period of time that a condition exists and the plant management is informed that if the recommended steps are <u>not</u> taken, it will add, for instance, $1.87 to the cost of each unit produced within the next 30 days.

The plant manager is also required to submit, in writing, the corrective action which he proposes to take to correct this condition within the next 48 hours and to forward his answer to the Quality Control Department. If the Quality Control Department does not receive an answer within 48 hours, or if the quality condition has not received suitable corrective action within 48 hours, the original report, along with the corrective action specified by the plant manager, is forwarded to top management.

Utilizing the capabilities and techniques of both Industrial engineering and Quality Control in a collective effort such as this, enables all of us to employ the proper action at the proper time to establish a sound hard-hitting organization that can offer to top management and to our customers, an honest and valuable service.

One of the most important phases of any program is often lost in the enthusiasm and drive that is required to carry it out. That particular phase is salesmanship. No person really accomplishes anything merely because he wants to. There is hidden somewhere in each one of us a deep, motivating force that pushes us forward to greater things. If we realized this fact, more accomplishment can be obtained than was ever dreamed of previously.

Quality Control, Industrial Engineering, Cost Controls and many other services are intangibles. They consist largely of systems, methods and procedures. Hard-hitting manufacturing people are sometimes too busy to reap the maximum benefits of these things unless they become a very real and active part of such a program.

This means that these intangibles must be sold all the way down the line ..... and up ..... in the greatest sales campaign we are able to muster. Accomplishing this, we, and the services which we render, will never fall behind the tangibles.

In this new age of automation, where industry is using the many modern techniques and methods which have been developed to increase efficiency and productivity, we, in the service organizations must do our share by collectively providing for industry an accurate evaluation of their capabilities and economics.

Our programs must be based on intelligent and logical facts, designed to offer a service where it is needed, rather than where it is convenient.

For many years, the question of applying statistical techniques to one area first and then enlarging it or installing a complete program in all areas and run the risk of spreading yourself too thin, appeared to be a matter of debate among the philosophers of our science. As many companies have discovered recently, this question becomes insignificant when a program based on common sense is followed. If we will use our thinking and planning to direct our efforts in the logical places where someone needs help, no matter how big or how small that area is, the question will answer itself.

Often the problem at hand will require merely a revision of procedure or inspection technique, while other times we may have to conjure many integrated systems. Whichever it is, we must remember that our service is only as valuable as the results which it produces. Common sense, as well as monetary cents can never be replaced by overly-ambitious statisticians. If we, who must often-times prove our efforts, abide by this rule, our efforts will not be in vain.

Visionary Product Engineering, new materials and incredible machines and tools can then, with our efforts, enable industry to attain even greater stature, by producing quality products at a cost which is compatible to our American economic future.

CONTROL CHARTS IN A PETROLEUM REFINERY LABORATORY

Charles R. Haag
Esso Standard Oil Company

The methods of modern statistics have found their way into the petroleum refinery testing laboratory only within the past five years. The introduction of these techniques follows the same path that was started by the mechanical and electrical industries almost 20 years ago. This path involves the work and problems which face the testing laboratory in any industry. Briefly, these problems are:

1) Inspecting the entering raw materials for quality
2) Providing data with which to run the manufacturing process
3) Examining the finished product for conformance to the specification.

In the Bayway Refinery, the raw materials are the crude oils and chemicals. The manufacturing process is the complex integration of distillation, catalytic cracking, polymerization, treating, blending, etc. The finished products are motor fuels, burner fuels, solvents and petrochemicals.

One of the main products of a laboratory in a petroleum refinery is data. We have put the statistical quality control chart to use in the laboratory and have improved the quality of the data on which we base our decisions as to how to run our business. In several areas, an improvement in precision was accomplished by statistically designed experiments.

This paper describes some of the problems and the results achieved through the use of statistical techniques.

The laboratory of a petroleum refinery usually provides 7-day round the clock service to the refinery as a whole. In Bayway, the data which comes from this service are the results of about 300 to 400 different testing methods ranging from the simple gravity and identification tests to the relatively new complex methods of instrumental chemistry. High precision is needed in the test methods used in petroleum refining since a large part of the day to day variability in the process streams is often due to the test method. Many of the processing units in a refinery need only a few operators and handle as much as a quarter of a million gallons of liquid feed each 24 hours and do this for many months in succession.

The introduction of the statistical control chart to improve the routine laboratory service was started in a modest way several years ago. It was apparent that in order to achieve any worthwhile improvement in our laboratory data, it was necessary to institute a short training program for the laboratory foremen since the responsibility for laboratory work quality rests with this group. A curriculum was designed to instruct these men in the theory, mechanics and necessary calculations to set up and maintain laboratory control charts.

The formal instruction consisted of four 1-1/2 hour sessions. The lecture presentations were pitched at the practical level, supplying

only the minimum of statistical theory. The outline of the course is as follows:

First Session: <u>Fundamental Concepts</u>
- a. demonstration with models and gadgets
- b. definition of "quality" and "statistics" implied in the title
- c. patterns of "normal" variation

Second Session: <u>Types of Variation</u>
- a. frequency distribution
- b. demonstration of kind of data used
- c. concept of "population" and sampling for an estimate of variation

Third Session: <u>The Control Chart for Variables</u>
- a. application in the laboratory
  1. standard sample
  2. for duplicate analyses (one example for each type)
- b. application in the plant
  1. demonstration with case history

Fourth Session: <u>The Control Chart (continued)</u>
- a. for moving averages and range
- b. discussion of mechanics for maintaining charts - illustrated with advanced case histories

The sessions were spaced two weeks apart to allow practice of the techniques between classes, since most of the examples used during class were drawn from current laboratory data. An intensified follow-up program was carried out after completion of the formal instruction. It consisted of technical assistance to compile data, set initial control limits, and provide a system of communication to the laboratory management on the progress of all active control charts. This intensified follow-up was in large part responsible for the continued interest and improved effectiveness of the foremen in translating the subjects of the lectures into practical applications.

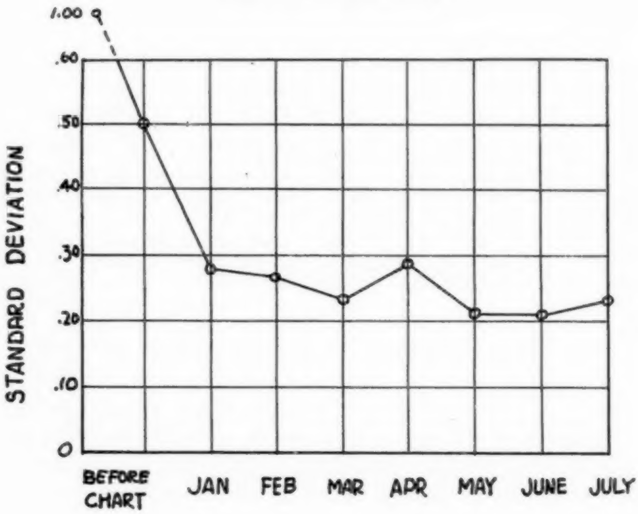
The results of the statistical control chart approach to the improvement of laboratory testing have been most gratifying. A good example is the important vapor pressure test for gasoline. This test and many others are described in "The Significance of Tests of Petroleum Products," a report by ASTM Committee D-2.

The A.S.T.M. Standard Method of Test for Vapor Pressure of Petroleum Products (Reid Method) (D 323-43) outlines the apparatus and procedure for the determination of the vapor pressure of volatile, nonviscous petroleum products. The importance of Reid Vapor Pressure to the customer is in the operation of his automobile. The specification for motor fuel quality calls for a maximum Reid Vapor Pressure as a safeguard against vapor lock in the fuel system. It has been found practicable to vary the maximum with the seasons, that is in the summer the Reid Vapor Pressure of the gasoline is lower than in the winter. The cold winter weather requires that increased volatility so that starting is easier. If the test method for measuring vapor pressure is too variable,

the motorist will experience poor car performance, also, the manufacturer will be at a disadvantage since he will not be able to incorporate the economic optimum quantity of the light naphthas in the gasoline.

The following chart shows the improvement in standard deviation in the Reid Vapor Pressure Test as a result of the control chart program to improve precision:

FIGURE I
REID VAPOR PRESSURE
Test Method Variability



If the test had not had the attention of a control chart it would have been necessary to run at least four tests at the same old precision to produce an average with the same size confidence interval. This is readily seen since the standard deviation of an average is smaller by the reciprocal of the square root of the number of individual tests in the average. Herein lies the economic incentive for the control chart. With more effective use of laboratory supervisory manpower, we have saved the price of the 3 extra tests or 3 times the cost of running just this one test.

It is important to distinguish between the precision of routine laboratory testing and the precision associated with just one operator or one chemist who might be running his own research program. In a refinery laboratory, there may be 75 to 100 skilled laboratory technicians across the three shifts. When a sample arrives for analysis, say a Reid Vapor Pressure determination, the laboratory foreman assigns it to any one of several technicians who may be available. The measure of the overall laboratory precision is achieved when the same sample is resubmitted on the next shift to another laboratory technician. The difference between these two independent measurements on the same sample is the familiar range of size n=2. This places one point on the range chart for the

test. The limits for the range and the appropriate factors to convert it into the standard deviation as shown in Fig. 1 are given in the A.S.T.M. "Manual on Statistical Quality Control of Materials."

It can be argued that a standard deviation obtained this way has in it pieces of variability due to shift and technician differences. This is desirable, however, since it is a practical kind of standard deviation and reflects the precision of the whole laboratory organization which is what other parts of a refinery organization look at. The foremen who have been trained by classroom and by practice to maintain the control charts spot some inconsistencies amongst themselves and iron out differences due to people and to shifts at their own level, which is just as it should be.

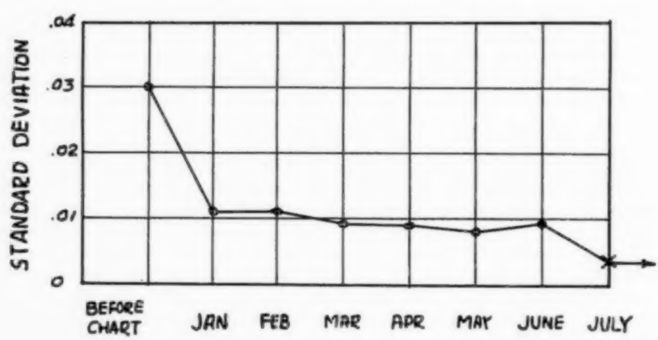
Occasionally, permanent samples are submitted on a blind basis to check the test method for accuracy. Frequently these samples are aliquots from reference standards carefully preserved for this purpose.

Another important test in the many that are being monitored for precision by the statistical control chart is that called the Carbon Residue Test. This test is important to the consumer of heating oil. It was originally developed by P. H. Conradson in 1912, and measures in part the tendency toward formation of carbonaceous deposits and residues in certain types of oil burning equipment. After over 40 years of use, it would be expected that this test technique would have been polished to perfection.

The control chart data indicating the precision of the test is shown in Figure 2.

The improvement in the standard deviation came immediately after control chart vigilance was placed on the test method. The last point which has been circled represents an improvement beyond that previously attained and was achieved by a planned experiment.

FIGURE 2
CARBON RESIDUE TEST



The test method for Carbon Residue has long been specified by the A.S.T.M. Standard Method (D189-41). The details of the test are written with good clarity. However, there are several places in the details of the test method where the technician operator can expand the precision of the test. Five of these places were examined at two levels, forming a 2^5 factorial experiment. Two of these factors were found to increase the variability by significant amounts. The first important factor was in the beginning of the test where the fuel oil is distilled to leave 10% of the original weight in the distillation flask. When this was done by weight instead of by volume, greater precision resulted. The second important factor came at the end of the test and involved firing a crucible at "cherry red." A technician's judgement as to what constitutes cherry red is not always reliable. The precision was further improved by substituting a high temperature furnace at constant temperature for his judgement as to what was a "cherry red" temperature.

The statistical control chart as a means of monitoring the quality of the data output of a refinery laboratory has eliminated a large part of the uncertainty in the minds of the people who use the data to run the process or decide on problems of product quality. About 90% of the applications made so far of this technique have shown considerable improvement in precision. Emphasis is placed on those tests of high economic value. Others are added as the amount of control chart testing is decreased for those already so monitored. In general, control chart testing accounts for about 10% of the total work load (for tests monitored with charts).

A bricklayer boss can look at a finished brick wall and see if it is plumb and straight. He tells by eye how well his masons are doing. Prior to the statistical control chart, our foremen could look at testing data and see that the samples were analyzed and reported on time to the production organization. He only had "feelings" in regard to precision. With the use of control charts to monitor the precision of testing and the introduction of occasional standardized samples to maintain accuracy, our control laboratory foremen can now tell by eye how well the technician and the test method are doing. This means our data are a workmanlike product - adequate to answer important questions about finished product quality.

SOME STATISTICAL PROBLEMS ENCOUNTERED IN INDUSTRIAL RESEARCH

W. S. Connor
Johnson & Johnson

Introduction. Industry is appreciative of statistical techniques as a management tool. This has been amply demonstrated in the area of statistical quality control, and is currently exhibited by interest in statistically planned experiments.

It is gratifying to find executives and their advisers relying on statistical methods to help solve problems. In this paper two such instances will be discussed.

The techniques used were not in the area of conventional statistical quality control, nor in the area of planned experiments. However, both problems concern quality, and because they are of general interest, it is profitable to consider them from the viewpoint of experiment design.

A storage problem. The first problem arose in an industrial engineering study conducted by Mr. T. J. Gorman. The study dealt with the effect of various storage conditions on the quality of a particular product. An objective was to determine suitable packaging and storage conditions.

The data available were the numbers of defectives out of 195 specimens of the product stored under each of five different storage conditions. The per cents defective for each of eight kinds of defect, together with the storage conditions, were as follows:

Per Cent Defective

Kind of Defect	1	2	3	4	5	6	7	8	Storage Condition
	14	9	97	17	94	5	8	9	1
	0	1	1	1	1	5	1	1	2
	4	5	6	5	7	9	6	1	3
	0	0	3	0	5	5	5	0	4
	3	2	34	1	34	12	58	6	5

	Storage Condition				
	1	2	3	4	5
Temp. ($^oF.$)	150	75	75	100	100
Time (Days)	6	14	120	7	90
Relative Humidity (%)	35	35	35	17	17

In two instances, the same set of specimens was observed at two different times, so in all, there were three distinct sets of 195 specimens.

It was thought that an equation of the form

$$p = b_o + b_1x_1 + b_2x_2 + b_3x_3$$

might fit the data, where

p = per cent defective
x_1 = temperature
x_2 = time
x_3 = per cent relative humidity

and the b's are regression coefficients.

Fitting the equation by Least Squares, it was found that the variations in the per cents defective were largely explained by variations in temperature, time, and per cent relative humidity. The per cent of the variation explained for each kind of defect was as follows:

Kind of Defect	1	2	3	4	5	6	7	8
Per cent Explained	99.9	99.9	99.4	99.3	97.3	88.6	66.5	60.0

Accordingly, it was felt that the fitted equations could be helpful in choosing a suitable package and in specifying desirable storage conditions.

The use of only five storage conditions does not provide much opportunity to evaluate the adequacy of the postulated linear relationships. However, the high per cents explained for defects 1 through 5 are persuasive.

The available storage conditions do not make for easy arithmetic in determining the regression coefficients. There are more tractable choices, as will be seen below.

There are important considerations which limit the storage conditions to those used. However, suppose that such limitations did not exist. Then the choice of storage conditions could fruitfully be considered in the realm of the design of experiments.

Suppose that the postulated linear relationship is known *a priori* to be adequate. Then the only statistical problem is that of determining the regression coefficients. This can be done using only four distinct storage conditions. Assume for each variable that the lower and upper limits which can conveniently be attained experimentally are, for example, as follows:

	Lower Limit	Upper Limit
Temp., °F.	75	150
Time, days	6	120
Rel. Hum., %	17	35

Then the following storage conditions may be used, with equal numbers of specimens at each condition:

Temp., °F.	Time, days	Rel. Hum., %
75	6	17
150	120	17
150	6	35
75	120	35

The arithmetic for the Least Squares fit is now very simple, and the regression coefficients b_1, b_2, and b_3 are determined as precisely as is possible for the conditions specified.

If it is not known a priori that the linear model is adequate, then the choice of storage conditions could allow for a test of this point. One possibility is to run the full 2^3 factorial in duplicate. Then the residual variation after the fit could be compared with the agreement among duplicates.

An inspection problem. The second instance concerns the comparison of inspection and quality levels at two different production centers. The approach developed by Mr. R. N. Brownlee is an unusually clever analytical attack on a familiar problem.

The product is produced at widely separated centers, M and S. Inspection is conducted at both centers, and there is the continuing problem of comparing quality and inspection levels, with the purpose of equalizing them. The program adopted involves certain exchanges of product for inspection purposes. These exchanges will be specified below.

Some notation is needed. Let

Q_{MS} = the true average quality of that portion of M's production which is sent to S for inspection.

Q_{MR} = the true average quality of that portion of M's production which is not sent to S, i.e., M's "regular" production,

and let Q_{SM} and Q_{SR} have similar meanings for S. In addition, for any of these Q's, let $(Q + I_M)$ denote the true average quality as found by M's inspection, so that

I_M = the systematic error in inspection for M,

and assign a similar meaning to I_S. Finally, let T denote the shift in any Q which is attributable to the travel of the product between centers.

Data reflecting quality were collected over a period of time, and it was found by Mr. Brownlee that they could be summarized in the following equations:

$$\begin{array}{lllllllll}
Q_{SS} & & & & + I_S & & & = & 94.4 \\
Q_{SS} & & & & & + I_M & + T & = & 94.1 \\
 & Q_{SR} & & & & + I_M & + T & = & 92.9 \\
 & & & Q_{MR} & & + I_M & + T & = & 92.3 \\
 & & Q_{MS} & & + I_S & & + T & = & 87.3 \\
 & & Q_{MS} & & & + I_M & & = & 91.1 \\
 & Q_{SR} & & & + I_S & & & = & 94.2\ ,
\end{array}$$

where the data on the right hand side are appropriate measures of quality.

Though there are seven equations in seven unknowns, the equations are not all independent, and it is necessary to impose one restriction on the constants in order to solve. A suitable restriction appeared to be

$$I_M + I_S = 0$$

Solving by Least Squares, it was found that the estimates of the constants and their standard deviations are as follows:

Constant	Estimated constant	Estimated Std. Dev.
Q_{SS}	95.40	.42
Q_{SR}	94.70	.42
Q_{MS}	90.35	.42
Q_{MR}	93.85	.65
I_S	-.75	.22
I_M	.75	.22
T	-2.30	.43

Because the standard deviations are estimated from only one degree of freedom, they are not very precise.

The utility of the approach is apparent. Its success lies in the careful formulation of the data as sums,or linear functions, of meaningful constants.

It is instructive to view the problem from the standpoint of experimental design. An aim would be to introduce more symmetry into the program.

This could be achieved, for example, by collecting data which would satisfy the following equations:

$$\begin{array}{llllllll}
Q_{SS} & & & & + I_S & & & = X_1 \\
Q_{SS} & & & & & + I_M & + T & = X_2 \\
& Q_{SR} & & & + I_S & & & = X_3 \\
& Q_{SR} & & & & + I_M & + T & = X_4 \\
& & Q_{MS} & & & + I_M & & = X_5 \\
& & Q_{MS} & & + I_S & & + T & = X_6 \\
& & & Q_{MR} & & + I_M & & = X_7 \\
& & & Q_{MR} & + I_S & & + T & = X_8 ,
\end{array}$$

where the X's are the measures of quality. Because of the symmetry, the Normal equations would be relatively easy to solve.

APPLICATION OF THE ANALYSIS OF VARIANCE TO PROBLEMS
IN METALLURGICAL RESEARCH

J. D. Hromi
Applied Research Laboratory
United States Steel Corporation

Introduction

The analysis of variance is a statistical technique developed about thirty years ago by R. A. Fisher to facilitate the analysis and interpretation of the data from experiments in agriculture and biology. Since many of the statistical designs of experiments used in agricultural and biological research have been adapted for use in engineering and the physical sciences, it is only natural that the analysis of variance procedure has become an important tool in metallurgical research.

In seeking a solution to a particular problem, the research worker first formulates hypotheses about the problem. Then the experimenter must gather observational data to either support or disprove his hypotheses. It is on this aspect of the scientific method that my colleague, S. Gilbert[1)], spoke at the last Annual Convention of the American Society for Quality Control. At this point, it may be well to review some of the principles set forth in Gilbert's paper on the design of experiments.

Experimental data must be collected according to some planned scheme in order to attain the objectives of the investigation as economically and efficiently as possible. Frequently, too little time and effort is devoted to the planning of the investigative procedure. As a result, the analysis and interpretation of the data often do not provide the experimenter with sound answers to specific questions. The experiment should be planned in detail. Some planners even suggest a written outline enumerating the proposals for the experiment. It is contended that an outline or check list facilitates the development of an efficient experimental program to provide data pertinent to the objectives under study. Rarely can the statistician analyze and interpret successfully the data resulting from experiments not previously designed to answer specific questions. Thus, to prevent costly expenditures of time and funds, one cannot emphasize too strongly that the time to think about statistical conclusions is during the planning stage of the experiment. If the design of the experiment is without fault, then the proper methods of interpretation should yield inferences with established confidence.

The problems confronting the research workers at the United States Steel Corporation's Applied Research Laboratory are generally those for which a controlled comparative type of experiment can be conducted. An experiment in which the investigator fixes the levels (given values or conditions) of the independent variables is almost always of the comparative type. As an example, the problem of making a quantitative evaluation of the effects of several process variables on the mechanical properties of a steel is typical of the programs conducted at the Laboratory. This experiment, which involves the evaluation of the effects of several factors or independent variables at two or more levels each, is a multiple factor experiment. The techniques of factorial experimentation were also developed by Fisher[2)] and his colleagues

in the science of agriculture, but are applicable to a major portion of the problems in metallurgical research.

The multiple factor experiment has several distinct advantages over the classical concept of experimentation in which all the independent variables but one are held constant. The factorial experiment has greater efficiency in that all the observations are used in drawing each conclusion. Therefore, each inference is made with increased precision. The factorial experiment also enables the research worker to evaluate interaction effects if they exist. The interaction measures the failure of the effect of one factor to be the same for all levels of another factor. If the effects of two factors, A and B, each at two levels are being studied, the simplest factorial design is employed. Each run can be denoted by a combination of the factors for a given level of each factor. For instance, the run in which factor A is at the high level and factor B is at the low level can be denoted by "a" on the basis that the presence of the letter in the combination denotes the high level of the factor and the absence of the letter from the combination denotes the low level of that factor. When both factors are at the low level, (1) symbolizes that particular run. Each run in the two-factor experiment is listed in Table I.

Table I

Run	Combination of Factors
1	(1)
2	a
3	b
4	ab

A combination of factors is often referred to as a "treatment combination". The four treatment combinations of the basic factorial design enable the experimenter to evaluate the main effects of A and B and the A x B interaction effect. The main effects are estimated by comparing the two runs made at the lower level of the factor with the two runs made at the higher level. As an example, the main effect of factor A is determined by comparing Runs 2 and 4 with Runs 1 and 3. The A x B interaction is determined by comparing the difference in the observations of Runs 1 and 2 with the difference in the observations of Runs 3 and 4. When no A x B interaction exists, the main effects are said to be additive. That is, the main effect of factor A is the same regardless of the level of factor B. Non-additivity prevails when the effect of factor A is dependent upon the value of factor B. If subsequent analysis reveals the non-existence of the A x B interaction, the factorial experiment is still more efficient than the classical experiment due to the "hidden replication", or built-in duplication.

The simplest factorial experiment is a special case of the general class of the 2^n factorial designs where the exponent n indicates the number of factors, each at two levels, to be studied. Thus the basic factorial experiment is more specifically referred to as a 2^2 experiment, and a three-factor experiment in which two levels of each factor are investigated is called a 2^3 factorial experiment.

The 2^3 factorial design enables the experimenter to obtain information on first, the main effects of three factors, A, B, and C; second,

the two-factor or first-order interaction effects; and third, the three-factor or second-order interactions. The treatment combinations used in the 2^3 experiment are listed in Table II.

Table II

Run	Treatment Combination
1	(1)
2	a
3	b
4	c
5	ab
6	ac
7	bc
8	abc

Treatment combination abc is the combination of factors in which each factor is at its upper level, but in treatment combination bc, only factors B and C are at their upper levels and factor A is at its lower level. As in the 2^2 experiment, each effect is estimated by comparing two sets of runs. In the 2^3 experiment, the data from the eight runs would provide a basis for arriving at seven conclusions on the effects produced by the variation of three factors. This can be illustrated by considering a particular example.

Recently, a program was initiated at the Applied Research Laboratory to determine whether the test results obtained from two testing machines are alike. Two operators determined the test values for six specimens from each of two materials on each of the two machines. Since the purpose of this example is to illustrate a method rather than to report specific information on the performance of the testing machines, it will suffice to refer to the three factors being studied as A, B, and C. The eight treatment combinations presented in Table II were repeated three times, and the eight runs of each replicate were completed in a random order. The observed values are coded and presented in Table III.

Table III

Treatment Combination	Coded Test Values
(1)	6, 6, 7
a	5, 4, 6
b	6, 7, 7
c	8, 9, 8
ab	3, 5, 5
ac	7, 7, 7
bc	9, 7, 8
abc	7, 6, 8

The seven differences or contrasts which estimate the effects of the factors may be determined in any one of three ways discussed hereafter.

Methods for Determining the Effects of Factors in a 2^n Experiment

One method for obtaining the seven conclusions from the 2^3 experiment is the placing of the experimental data into the appropriate four cells of a series of "2 x 2" tables. Each of the four cells in any two-way table contains the average of the six runs which fit the cell specification. The following two-way table is obtained from the data presented in Table III.

		Factor A		
		Low	High	Average
	Low	7.3	6.0	6.7
Factor B	High	7.3	5.7	6.5
	Average	7.3	5.9	

The effect of changing factor A from its low level to its high level is estimated by the difference in the column averages. Similarly, the effect of factor B is estimated by the difference in the row averages. The A x B interaction effect is estimated by the differences between the diagonal averages. The remaining conclusions are drawn from the two-way tables for factors A and C, for factors B and C, and for factors B and C at each level of A. The A x B x C interaction effect is estimated from the data combined in the latter two tables. The seven estimates obtained from the two-way tables are listed below.

Effect	Estimated Value
A	-1.4
B	-0.2
C	2.0
A x B	-0.2
A x C	0.3
B x C	0
A x B x C	0.3

Another technique for obtaining the seven conclusions from the 2^3 experiment is illustrated in Table IV.

Table IV

	Effect						
Treatment	A	B	C	A x B	A x C	B x C	A x B x C
(1)	-	-	-	+	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	-	+	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	+	-	+	-	-
bc	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+

To evaluate the effects shown in Table IV, the sum of the observed values for those treatments indicated by a minus sign is subtracted from the sum of the observed values indicated by a plus sign. Dividing the difference by twelve gives the average effect or contrast. For example, the estimate of the average B x C effect is $1/2 \left[(6 + 6 + 7 + 5 + 4 + 6 + 9 + 7 + 8 + 7 + 6 + 8) - (6 + 7 + 7 + 3 + 5 + 5 + 8 + 9 + 8 + 7 + 7 + 7) \right] = 0$. This value checks the B x C interaction as determined from the two-way table for factors B and C. Table IV is quite simple to construct. For the main effect of any factor, a plus sign is placed opposite each treatment combination in which that factor appears at the high level. The signs under each interaction are the algebraic products of the signs under the corresponding main effects. The construction of Table IV is covered in greater detail in statistical textbooks by Kempthorne[3)], Cochran and Cox[4)], and Davies[5)].

The third systematic method of obtaining the seven contrasts from the 2^3 experiment is attributed to Yates. The device used in completing the analysis of the experimental data is presented in Table V.

Table V

Treatment	Observed Value	Column 1	Column 2	Column 3	Effect
(1)	19	34	67	158	Total
a	15	33	91	-18	A
b	20	46	-11	-2	B
ab	13	45	-7	-2	AB
c	25	-4	-1	24	C
ac	21	-7	-1	4	AC
bc	24	-4	-3	0	BC
abc	21	-3	1	4	ABC

The number of columns for carrying out the numerical process is three. Generally, the number of columns is determined by the exponent n in the 2^n designation. The first half of the entries in Column 1 is obtained by summing succeeding pairs of the observed values and the second half of the entries in Column 1 is obtained by determining the differences for these same pairs of observed values. The first number of the pair is always subtracted from the second number. Columns 2 and 3 are constructed in the same manner as Column 1, the data in Column 1 being used for constructing Column 2 and the data in Column 2 being used for constructing Column 3. The total effects are read directly from the n^{th} column and the average effects are then obtained by multiplying the total effects by $1/r2^{n-1}$ where r is the number of times the experiment is repeated. In the aforementioned 2^3 experiment, the average effects are one-twelfth of the total effects. Values obtained from Yates' calculation checked those previously determined.

The three methods for obtaining (2^3-1) conclusions from a 2^3 experiment can be conveniently extended so that (2^n-1) conclusions may be drawn from a 2^n experiment. However, the first two systems for establishing the effects of the factor become somewhat complex as the value of n in the 2^n factorial design increases. Therefore, Yates' method seems to be simpler for handling the data from the larger experiments. An additional advantage that the Yates' method has over the other two methods is the column check. The calculations for each column can be verified before proceeding to the following column. The details of

Yates' method and the associated check are found in a textbook by Davies[5].

Experimental Designs Other Than the 2^n Factorial Design

The complete investigation of the effects produced by the variation of n factors from one level to another, in the context of all possible combinations of the other (n-1) factors, would require 2^n runs. The number of main effects and of low-order interactions is listed for several 2^n factorial designs in Table VI, which also demonstrates the rapid increase in the scope of the experiment with each increase in n.

Table VI*

Effects	Number of Factors, n						
	2	3	4	5	6	7	8
Main	2	3	4	5	6	7	8
Two-factor interactions	1	3	6	10	15	21	28
Three-factor interactions		1	4	10	20	35	56
Four-factor interactions			1	5	15	35	70
Total number of runs	4	8	16	32	64	128	256

*A more extensive table is given by K. A. Brownlee[6].

It is obvious that from the standpoint of economy and practicability, the number of experimental runs becomes prohibitive. Then too, the larger 2^n experiments provide the experimenter with information on high-order interactions which are likely to be negligible and of no interest to him. Therefore, it is desirable to plan a smaller experiment. A smaller and more practical experiment is obtained when the research worker selects a fraction of the runs required for a full factorial experiment. The fractional design is not obtained by choosing just any set of runs. The correct set of runs to use in a fractional replicate of a factorial design is determined by methods described in statistical textbooks by Kempthorne[3], Cochran and Cox[4], and Brownlee[6]. The three methods of analysis already discussed are also applicable to data obtained from fractional factorial experiments.

The experimental designs used in the Applied Research Laboratory are not limited to those discussed here. Several factors, each at two or more levels, may cause variations in the values of the dependent variable when the factors are changed from level to level. For example, our Laboratory desired to determine whether the method of supporting a certain type of test specimen had any effect on the test result. The effect was evaluated for three different steels, each in five different stress-relief annealed conditions, at each of two testing temperatures. The experiment was planned and performed as a $2^2 \times 3 \times 5$ experiment; two factors were investigated at two levels each, one at three levels, and one at five levels. The methods for systematic analysis of the experimental data discussed heretofore do not apply in the strict sense to the $2^2 \times 3 \times 5$ experiment. However, the two-way tables can be modified in order to prepare the experimental data for use in the analysis of variance.

The preceding discussion has been concerned with the analyses of laboratory data obtained according to sound experimental plans, which

analyses give the experimenter estimates of the effects of the factors he is investigating. Confronted with the estimates, the experimenter must judge whether the effects are real. This is the basis of statistical inference.

Statistical Inference and the Analysis of Variance

Statistical inference is the process of generalizing from particular results. Thus laboratory experiments are performed on a small scale in order to predict behavior in a large-scale process. Often laboratory experiments do not make the same set of predictions when repeated. For this reason, it is the problem of statistical inference to provide with some predetermined degree of certainty general conclusions from the experimental data. The measure of variation is expressed as the variance, and the measure of uncertainty is expressed as a probability.

Variance is the average of the squares of the deviations of the observed values from their mean. The formula for calculating the variance is:

$$\hat{\sigma}^2 = \frac{\sum (X - \bar{X})^2}{n}$$

where $\hat{\sigma}^2$ is an estimate of the true variance, X is an individual value, $\bar{X}$ is the average of the set of values, and n is the number of values in the set. For small groups of data as obtained in laboratory experiments, the variance calculated from the above formula is biased—it is ordinarily too small[7]. Thus, for small sets of data, it is better to use the formula

$$s^2 = \frac{\sum (X - \bar{X})^2}{n-1}$$

where s^2 is also an estimate of the true variance, σ^2. This second formula gives an unbiased estimate of the variance of small sets of data. The square root of the variance is called the standard deviation and is denoted by the symbol s. The variance is a widely used characteristic of experimental data because it has maximum efficiency as a measure of variability. It also has the property of being additive; that is, separate variances may be summed. Conversely, the partitioning of the total variance into its component variances, each attributable to a particular experimental factor or combination of factors, is accomplished by the analysis of variance.

The analysis of variance enables the experimenter to test for real, or significant, differences among two or more means. It has been established that when the means of subsets of data are significantly different, the variance of the combined sets is much larger than the variances of the separate sets. The analysis of variance also enables the research worker to detect and estimate components of random variation. Problems of this nature, however, are beyond the scope of this paper.

Since the computational procedure and the mechanics of the statistical tests of significance are the same for both uses of the

analysis of variance, a standard form for completing the analysis is employed. This form is given in Table VII for the 2^3 factorial experiment.

Table VII

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	Variance Ratio
A		1		
B		1		
C		1		
A x B		1		
A x C		1		
B x C		1		
A x B x C		1		
Total		7		

The column at the left shows how the total variance is partitioned into portions attributable to each of the seven sources of variation. The numbers to be entered in the second column are the sums of the squares of the deviations from the appropriate averages. Methods for calculating the sums of squares are demonstrated in the textbook by Davies[5]. The total sum of squares is obtained by squaring the deviations of each of the observed values from the grand average. The degrees of freedom are determined from the number of means that can be varied freely when the grand average is established. For example, if the mean of the treatments with A at the low level is free to take on any value, the mean of the treatments with A at the high level is fixed; its value is restricted by the established grand average. Thus in the 2^3 experiment, only one degree of freedom is associated with each variance. The total degrees of freedom are obtained by reasoning that the grand average is fixed and that one of the eight results is thereby controlled, leaving seven degrees of freedom. The mean squares (or variances) are the mean-square deviations obtained by dividing each sum of squares by its associated number of degrees of freedom. The F ratio, or variance ratio, is the quotient of the mean square for some factorial effect divided by the mean square for experimental error. The calculated variance ratios enable the experimenter to judge with some predetermined probability of error whether a factorial effect is real.

Mathematicians have tabulated variance ratios* that are exceeded with a defined probability when the mean squares being compared are not different. For example, suppose the degrees of freedom of the mean square in the numerator of the variance ratio is 5 and that of the mean square in the denominator is 10, then a variance ratio of 3.3 or greater may be exceeded one time in twenty when the two variances are really the same. Therefore, if the experimenter judges an effect as real on the basis of the variance ratio, 3.3, he can be wrong one time in twenty.

---

*Variance-ratio tables are found in most modern textbooks on statistics.

The aforementioned principles were used recently in evaluating the effects of chemical composition on the physical characteristics of a particular grade of steel. It was decided to investigate the effects of six factors, at two levels each, in one experiment. To be general, the factors are referred to here as A, B, C, D, E, and F. The laboratory runs were those treatment combinations which constituted a 1/2 replicate of a 2^6 factorial design, Table VIII.

Table VIII

Run	Treatment	Run	Treatment
1	(1)	17	af
2	ab	18	bf
3	ac	19	cf
4	bc	20	abcf
5	ad	21	df
6	bd	22	abdf
7	ad	23	acdf
8	abcd	24	bcdf
9	ae	25	ef
10	be	26	abef
11	ce	27	acef
12	abce	28	bcef
13	de	29	adef
14	abde	30	bdef
15	acde	31	cdef
16	bcde	32	abcdef

Yates' method was used for the analysis of the experimental data. The mean squares listed in the analysis of variance table, Table IX, were obtained by squaring the appropriate contrasts from Yates' calculation and then dividing each square by the total number of observations. In the replicate of the 2^6 factorial design, the estimates of the main effects are masked by, or confounded with, the five-factor interaction effects and the estimates of the two-factor interactions are confounded with the four-factor interaction effects. This fractional factorial design can be used only if it is reasonable to assume that the high-order interactions are negligible. Since the three-factor interactions are confounded in pairs, it is impossible to test any three-factor interaction for significance. The sum of squares of the three-factor interactions confounded in pairs is the residual sum of squares. A variance ratio was determined for the mean square of each effect as judged against the residual or error variance. Factors A, D, and F and interactions A x D, A x E, and C x F produced in this experiment effects which were judged as real effects at the significance levels indicated in Table IX. Choosing the significance level should be a part of the planning; the experimenter decides in advance the frequency with which he can tolerate erroneous decisions.

Each analysis of variance is based on a mathematical model which should be selected in the planning stage of each experiment. In addition, certain assumptions underlying the analysis of variance[8,9] must be fulfilled if statistical inferences are to be made from the experimental data. Briefly, it is assumed that the numbers are observed values of random variables which are normally distributed with common variance and are not dependent on each other. Detailed discussion

regarding the mathematical model and the assumptions underlying the analysis of variance is not within the scope of this presentation.

Table IX

Source of Variation	Sum of Squares*	Degrees of Freedom	Mean Squares	Variance Ratio
A		1	193.06	18.20xx
B		1	6.67	0.63
C		1	36.13	3.41
D		1	64.41	6.07x
E		1	51.51	4.85
F		1	25673.78	2419.77xxx
A x B		1	12.50	1.18
A x C		1	15.40	1.46
A x D		1	67.28	6.34x
A x E		1	131.22	12.37xx
A x F		1	5.28	0.50
B x C		1	0.10	0.01
B x D		1	0.02	0.00
B x E		1	1.28	0.12
B x F		1	20.16	1.90
C x D		1	3.25	0.31
C x E		1	18.30	1.72
C x F		1	74.42	7.01x
D x E		1	0.18	0.02
D x F		1	12.75	1.20
E x F		1	1.20	0.11
Residual	106.10	10	10.61	
Total	26495.00	31		

*Since all but the residual effect are determined with one degree of freedom, the sums of squares and the mean squares are the same.

x denotes significance at the 0.05 probability level.
xx denotes significance at the 0.01 probability level.
xxx denotes significance at the 0.001 probability level.

Summary

Balanced factorial experiments are recommended because they permit conclusions of maximum generality, since all the data obtained are used in drawing each conclusion. Any one of three systematic methods of analysis—the series of two-way tables, the grouping of observed data according to an arrangement of plus and minus signs, or Yates' calculation—are used to facilitate the evaluation of the effects of the various factors. Finally, the analysis of variance is recommended for partitioning the total variance of a set of data into its component variances attributable to each of the factors or combinations of factors being studied. By comparing each of the component variances with the estimated error variance, judgment as to whether the observed effects are real is made with a predetermined degree of confidence.

References

1) S. Gilbert, "Statistical Design of Experiments in Metallurgical Research", ASQC Convention Transactions, 1954.

2) R. A. Fisher, The Design of Experiments, Hafner Publishing Company, Inc., New York, 1952.

3) O. Kempthorne, Design and Analysis of Experiments, John Wiley and Sons, New York, 1952.

4) W. A. Cochran and G. M. Cox, Experimental Design, John Wiley and Sons, New York, 1950.

5) O. L. Davies, The Design and Analysis of Industrial Experiments, Hafner Publishing Company, Inc., New York, 1954.

6) K. A. Brownlee, Industrial Experimentation, Chemical Publishing Company, New York, 1950.

7) G. W. Snedecor, Statistical Methods, Iowa State College Press, Ames, Iowa, 1950.

8) C. Eisenhart, "The Assumptions Underlying the Analysis of Variance", Biometrics, Vol 3, No. 1, March 1947, pp 1-21.

9) W. G. Cochran, "Some Consequences When the Assumptions for the Analysis of Variance Are Not Satisfied", Biometrics, Vol 3, No. 1, March 1947, pp 22-38.

10) R. L. Anderson and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill Book Co., Inc., New York, 1952.

11) D. J. Finney, "The Fractional Replication of Factorial Experiments", Annals of Eugenics, 12, pp 291-301, 1945.

12) D. J. Finney, "Recent Developments in the Design of Field Experiments: Part III Fractional Replication", Journal of Agricultural Science, 36, pp 184-191, 1946.

13) K. A. Brownlee, B. K. Kelley, and P. K. Loraine, "Fractional Replication Arrangements for Factorial Experiments With Factors at Two Levels", Biometrika, Vol XXV, Parts III and IV, December 1948.

14) A. L. Davies and W. A. Hay, "The Construction and Uses of Fractional Factorial Designs in Industrial Research", Biometrics, Vol 6, No. 3, September 1950, pp 233-249.

161 W Wio